Mod 7: Portfolio Milestone

Final Research Paper

Wilson H. Chan

Colorado State University Global

MIS581: Capstone – Business Intelligence and Data Analytics

Dr. Justin Bateh

April 9th, 2023

Table of Contents

Abstract

This research project aims to analyze the management of hospital capacity and vaccine distribution in the United States in response to the COVID-19 pandemic. The study will focus on the datasets "*COVID-19 Reported Patient Impact and Hospital Capacity by Facility*" and "*COVID-19 Vaccinations in the United States, County*," supplied by the US Department of Health and Human Services and the Centers for Disease Control and Prevention, respectively. The study will determine how hospitals managed capacity during the pandemic and how vaccines were distributed within the US. This study found significant differences in inpatient bed usage rate (IBUR), hospitalized adult COVID patients, and vaccination rate per state. This provides valuable insight into the response of the US healthcare system to the pandemic and may inform future pandemic management strategies. The US will need to provide additional resources to states with the high IBUR and low vaccination rates.

Introduction and Background

The COVID-19 pandemic has spread globally and has had a significant impact on public health, economies, and societies around the world. According to Jensen (2020), the virus originated in “Wuhan City, Hubei Province of China. The disease has spread to several other countries, including the United States. As of March 2, 2020, tens of thousands of people have been infected and over 2,500 have died. Both the WHO and the U.S. Centers for Disease Control and Prevention (CDC) post frequent updates on the outbreak”(p. 1). The virus had previously spread at such an alarming rate that hospital systems did not have the resources to manage the influx of affected patients. “The SARS-CoV-2 (COVID-19) pandemic has placed unprecedented demands on entire health systems and driven them to their capacity, so that health care professionals have been confronted with the difficult problem of ensuring appropriate staffing and resources to a high number of critically ill patients.”(Brüggemann et al., 2021, p. 1). The current state of the pandemic has become more manageable with the introduction of multiple vaccines and boosters, but the virus has caused lasting effects. After the introduction of the vaccine, hospital occupancy drastically decreased when compared to the height of the pandemic which is a major sign of the vaccine’s effectiveness. Researchers continuously track the spread of the disease and provide guidance for combating the virus as new treatment methods are developed.

Goals and Objectives

The goal of this research project is to determine how resources are managed within the US in response to the COVID-19 pandemic. “As the coronavirus (COVID-19) pandemic continues, government officials around the world have raised questions and concerns about how to ensure that their countries have adequate access to, and supplies of, medical goods, including personal protective equipment (e.g., masks, gloves, garments), medical equipment (e.g., ventilators), and pharmaceuticals”(Jensen, 2020, p. 77). The resources that will be covered by this research project include hospital capacity and vaccines. Resource distribution will be determined at the state level where possible.

The strategy to achieving the goal of this research project will be first to find appropriate data sources related to managing COVID-19 supplies. The scope of this project is narrowed to the management of COVID-19 within the US. This may also reflect the impact of the virus within certain regions within the US. In the earlier states of the pandemic, bedspace was often strained to the point where temporary facilities were required to accommodate more patients. According to Singh (2021), “the COVID-19 epidemic has required the significant capacity building to accommodate the increased number of critically ill patients. At the peak of the pandemic, many countries were forced to resort to the building of temporary structures to house critically ill patients, to help tide over the crisis”(p. 9). The US government provides a data repository on HealthData.gov where healthcare related data can be found. The website provides many datasets that focus primarily on COVID-19. Information regarding the virus is captured from its first case to its current state. The dataset “*COVID-19 Reported Patient Impact and Hospital Capacity by Facility*” supplied by US Department of Health and Human Services

(2023) will be analyzed to determine the fluctuation of hospital capacity over time. The dataset provides valuable information related to the goal of this research project. Some key variables contained in this dataset includes bed space usage. This can be compared to the total number of COVID-19 positive patients at any given time. The dataset also sorts information by the state level, so additional granular analysis can be made.

The deployment of the COVID-19 vaccine was heavily monitored in the early stages of the pandemic. Only high-risk groups were administered the vaccine. According to Shukla et al. (2022), “CVS Health was selected by Operation Warp Speed as a Federal Pharmacy Partner for vaccination and tasked with the vaccination of patients and healthcare workers in the US. The vaccines used in this initiative received emergency use authorization in December 2020 (Pfizer on December 21; Moderna on December 28)”(p. 734). The focus has now shifted to vaccinating the majority of the public. The dataset *“COVID-19 Vaccinations in the United States, County”* supplied by Centers for Disease Control and Prevention (2023) can be used to determine the distribution of vaccines within the United States. Some of the main variables in this dataset include the county vaccination percentage and total number of vaccines administered.

Overview of Study

The study aims to investigate how resources, including hospital capacity and vaccines, have been managed in response to the COVID-19 pandemic in the US. The goal of the study is to determine if the US has been able to manage these resources effectively to combat COVID-19. The study will analyze two datasets, namely the "*COVID-19 Reported Patient Impact and Hospital Capacity by Facility*" dataset and the "*COVID-19 Vaccinations in the United States,* *County*" dataset. The former dataset will be used to assess how hospital capacity was managed during the pandemic, while the latter will be used to determine the distribution of vaccines within the US. The research question to be addressed is whether the US was able to properly manage hospital capacity and vaccines to combat COVID-19.

Research Question

The main question that needs to be addressed is the following: Is the US able to properly manage resources (hospital capacity and vaccines) to combat COVID-19? This question will be analyzed by using the two aforementioned datasets. The dataset "*COVID-19 Reported Patient Impact and Hospital Capacity by Facility*" supplied by the US Department of Health and Human Services (2023) will be used to answer how hospital capacity was managed during the pandemic. The rapid spread of the virus caused an overwhelming strain on hospital capacity and personnel usage. The analysis of this dataset will provide clarity on how hospitals managed capacity to ensure that critical care was provided to those in need.

The dataset *"COVID-19 Vaccinations in the United States, County"* supplied by the Centers for Disease Control and Prevention (2023) will be used to determine the distribution of vaccines within the United States. The US administration tracked the vaccination rate per county and state within this dataset. The data is also organized by age group, so more granular analysis may be potentially performed per age group.

Hypotheses

The scope of this project will be narrowed down to the top 4 largest states by population according to the US census. The top 4 populated states include California, Texas, Florida, and New York. All variables identified in the following hypotheses are also time dependent. Therefore, time series regression can visualize each variable. “The distinguishing aspect of time series regression models is the common presence of serial dependence —a correspondence of values at different points in time— which does not typically arise in cross-sectional data. In time series regressions, there is a sense in which the order of the observations matters”(Darity, 2008, p. 371). Predictions based on time based regression coupled with hypothesis testing can guide future decision making.

First Hypothesis

The first hypothesis questions if inpatient bed usage rate is statistically different among the top 4 populated states. According to Alkarkhi (2021), “Hypothesis testing is a statistical technique used to make judgements regarding claims or statements related to populations. The opinion and definition of hypothesis testing with the general procedure of conducting the analysis are clarified before being applied to environmental data”(p. 2). The variables within the “*COVID-19 Reported Patient Impact and Hospital Capacity by Facility*” dataset will be used for this analysis. Inpatient bed usage rate (IBUR) will be calculated by using the formula:

$$IBUR = Inpatient\_beds\_used\_7\_day\_sum \div Inpatient\_beds\_7\_day\_sum$$

<u>Null Hypothesis (H_0)</u>: There is no difference between each state’s average IBUR.

<u>Alternate Hypothesis (H_a)</u>: At least one state’s average IBUR is statistically different than other states’ average IBUR.

The hypotheses can be summarized using the following equations where μ represents the average IBUR for each State:

$$H_0: \mu_{California} = \mu_{Texas} = \mu_{Florida} = \mu_{New\ York}$$

$$H_a: \mu_{California} \neq \mu_{Texas} \neq \mu_{Florida} \neq \mu_{New\ York}$$

Second Hypothesis

The second hypothesis aims to identify if there is a statistical difference in the number of hospitalized adult COVID patients within top 4 populated states. The variable total_adult_patients_hospitalized_confirmed_covid_7_day_sum within the "*COVID-19 Reported Patient Impact and Hospital Capacity by Facility*" dataset will be used for this analysis. This variable identifies the average number of confirmed adult covid patients within the hospital per week.

Null Hypothesis (H_0): There is no difference between each state's average number of adult COVID-19 patients.

Alternate Hypothesis (H_a): At least one state's average number of adult COVID-19 patients is statistically different than other states' average adult COVID-19 patients.

The hypotheses can be summarized using the following equations where μ represents the average number of adult COVID-19 patients for each state:

$$H_0: \mu_{California} = \mu_{Texas} = \mu_{Florida} = \mu_{New\ York}$$

$$H_a: \mu_{California} \neq \mu_{Texas} \neq \mu_{Florida} \neq \mu_{New\ York}$$

Third Hypothesis

The third hypothesis aims to identify if there is a statistical difference in vaccination rate among the top 4 populated states. The variable Series_Complete_Yes within the *"COVID-19 Vaccinations in the United States, County"* determines the total number of individuals that have been administered both doses of the COVID-19 vaccine. The variable Census2019 provides the number of people residing within each county. Vaccination rate can be determined using the following formula:

$$\boxed{\text{Vaccination rate} = \text{Series\_Complete\_Yes} \div \text{Census2019}}$$

<u>Null Hypothesis (H0):</u> There is no difference between each state's average vaccination rate.

<u>Alternate Hypothesis (Ha):</u> At least one state's average vaccination rate is statistically different than other states' vaccination rates.

The hypotheses can be summarized using the following equations where μ represents the average vaccination rate of each state:

$$H_0: \mu_{California} = \mu_{Texas} = \mu_{Florida} = \mu_{New\ York}$$

$$H_a: \mu_{California} \neq \mu_{Texas} \neq \mu_{Florida} \neq \mu_{New\ York}$$

Literature Review

For this project, the origin of COVID-19 and its effect on the global population is examined by Jensen in his article *COVID-19: Overview, Economic Implications and Federal Response.* His commentary provides substantial background information on the detrimental effects of COVID-19. The economic implications of the pandemic have been significant. Many countries experienced business closures, job losses, and reduced consumer spending. The pandemic has had a particularly severe impact on industries such as travel, hospitality, and entertainment. The federal response by the US to the pandemic has included several measures aimed at slowing the spread of the virus. This background information is highly relevant to this research project, however, the scope of this project only involves US resource management such as hospital capacity and vaccines.

The United States' management of hospital capacity during the COVID-19 pandemic is a focus of this research project. An abundant amount of previous research has been performed on this topic by scholars and epidemiologists. Brüggemann et al. provides an in-depth analysis of hospital resource management during the course of the COVID-19 pandemic in their article *Decision Support Tool for Hospital Resource Allocation During the COVID-19 Pandemic*. Another article written by Singh titled *Challenges faced in establishing a dedicated 250 bed COVID-19 intensive care unit in a temporary structure* provides additional clarity on resource allocation. Both articles provide background information on how hospital capacity was overwhelmed by the COVID-19 pandemic. They also provide commentary on the methods used to increase capacity to address these concerns such as implementing additional bedspace to manage additional COVID-19 patients. This research project attempts to identify which states within the US requires additional capacity based upon analyzing data supplied by the US federal government. These article provides guidance on how to improve capacity after areas in need are identified.

Another focus of this research project is the allocation of vaccines to combat the virus. Shukla et al. provides commentary regarding vaccine production, allocation, distribution, and its administration in their article *Optimizing vaccine distribution via mobile clinics: a case study on*

COVID-19 vaccine distribution to long-term care facilities. The vaccine was first distributed to health care workers and elderly individuals who were at high risk of exhibiting complications due to the virus. The vaccine has now been steadily administered to all risk groups. However, the vaccination rate varies across different states and regions, with some areas having higher vaccination rates than others. This project will attempt to determine which states require additional support from the federal government to increase completed vaccinations.

Two datasets were identified that individually tracked hospital resources and vaccine distribution. Both datasets were supplied by the healthdata.gov website which acts as a health related data repository managed by the US federal government. Alkarkhi provides a thorough transcript for understanding how to perform hypothesis testing is his article *Introduction to statistical hypothesis testing*. Null and alternate hypotheses were developed for the research questions involving the datasets. In most cases, a one-way analysis of variance (ANOVA) will be used to compare multiple groups in the datasets. Scheff provides an extensive explanation of statistical analyses including ANOVA in his article *One-Way Analysis of Variance*. Time series regression can also be used predict and visualize future resource allocation. Darity provides clarity on the use of time series regression models in his article *Time Series Regression*. Underlying trends or season patterns can be discovered using this technique.

Research Design

This project will involve two datasets. The dataset "*COVID-19 Reported Patient Impact and Hospital Capacity by Facility*" supplied by US Department of Health and Human Services (2023) contains multiple variables that will be analyzed to determine capacity management within hospitals. This dataset will be referred to as the Hospital Capacity dataset. The dataset

"COVID-19 Vaccinations in the United States, County" supplied by Centers for Disease Control and Prevention (2023) will be used to determine the distribution of vaccines within the United States. This dataset will be address as the Vaccinations dataset. The participants in both these datasets include the general population within the United States.

Methodology

Both datasets contain empirical data collected during the existence of COVID-19. Therefore, quantitative methods will be applied to both datasets. Data will be aggregated and analyzed using statistical methods. Quantitative research methodologies will identify if there are statistically significant results that can be generalized to other populations. Key variables applicable to the goal of this research project within both datasets are listed below.

The Hospital Capacity dataset is rather large with 128 column variables and 752352 records. Some key variables contained in this dataset are the following:

1) the total number of inpatient beds used per week

   (column: *inpatient_beds_used_ 7_day_sum)*

2) total number of beds available

   (column: *total_beds_7_day_sum*)

3) the total number of confirmed adult cases of COVID-19 within the hospital

   (column: *total_adult_patients_hospitalized_confirmed_covid_7_day_sum*)

4) the total number of confirmed pediatric cases of COVID-19 within the hospital

   (column: *total_pediatric_patients_hospitalized_confirmed_coivd_7_day_sum*)

These variables will be compared to determine if there is enough bed space within a hospital to accommodate the total number of COVID-19 positive patients within that hospital at any given time. The dataset also sorts information by the state level, so additional group analysis can be made.

The Vaccinations dataset contains 80 column variables and over 1 million records. Some key variables contained in this dataset are the following:

1) total series of vaccines [dose 1 and dose 2] completed

   (column: *Series_Completed_Yes*)

2) total population according to the 2019 Census

   (column: *Census2019*)

These variables can be compared to determine the vaccination rate of the total population within a given county. Since the dataset also groups the counties by the state level, the overall vaccination rate of each state can be determined.

Methods

The Hospital Capacity dataset and Vaccinations Dataset both contain numerical information. The information is less prone to bias due to the direct data collection methods mandated by government agencies. Hypothesis testing will be conducted for each set of hypotheses using an analysis of variance test to determine if there is a significant difference among all groups. According to Scheff (2016), “When comparing more than two groups it is necessary to use a statistical technique called analysis of variance (ANOVA). What this statistic does is examine the amount of variance in the dependent variable and try to determine from where that variance is coming”(p. 103). The test can be performed in SAS Studio by using the One-Way ANOVA task. Finding a statistical difference where a state has a high inpatient bed usage ratios indicates that officials should focus on increasing bed space capacity for that particular state. If there is a statistical difference where a state has a larger number of COVID-19 patients indicates that officials should provide additional resources to cover those patients within that state. If there is a statistical difference in vaccination rate where a state has a low vaccination rate, officials should aim to provide more vaccination resources and more vaccination information within those areas.

The previously identified data sources will be analyzed. Data analysis methods will be performed using SAS Studio. Descriptive analytics techniques such as summary statistics and distribution models will be developed. The primary goal of this analysis is to determine which states require the most resources. This includes metrics such as highest bed space capacity and highest number COVID patients. The states with the highest number of vaccinations will also be identified. After the preliminary review of the data is completed, data modeling and visualizations can be created for each dataset. Time series regression can be used to track the fluctuation of hospital capacity over time. If the model fit is appropriate, it may also be used to predict the need for increased capacity. Time series regression can also be used track the number of vaccines administered per state. Data visualizations can indicate which states are in need of additional vaccines via bar graphs and pie charts.

Limitations

Although the datasets presented within this research project are in-depth, there is little context behind the data collection methods used by government agencies. It is assumed that they follow regulatory standards so the data is accurate. However, it was also difficult to identify COVID-19 positive cases prior to development of advanced testing kits. Data prior to the development of these kits may be skewed as COVID-19 cases were misidentified. The data is also aggregated within both datasets, so it is difficult to distinguish where exactly data is collected. Granular details may be missing due to the aggregated nature of the datasets. The datasets do not reflect demographic disparities as well.

Ethical Considerations

Ethical concerns surrounding both datasets include the use of patient information. Both datasets do not have specific unique patient identifying information, so patient privacy is maintained. The data is also aggregated with almost no chance of revealing individual patients. However, the collection agencies that supply these datasets must have gained authorization from patients to use this data. Patient information contained within hospital systems is highly controlled and subjected to federal regulations. Researchers can identify trends and patterns associated with COVID-19 through the analysis of this patient data. This leads to evidence based decision making that will promote population health.

Study Findings

Hospital Capacity Dataset Descriptive Statistics

Figure 1: *CA Summary Statistics*

state=CA

Variable	Mean	Std Dev	Minimum	Maximum	N
InpatientBeds7DaySum	1185.17	1064.77	0	8206.00	53038
InpatientBedsUsed7DaySum	885.7927712	856.5897716	0	6379.00	53038
IBUR	0.7122544	0.2285018	0	1.2307692	52813
total_adult_patients_hospitalize	89.6623996	163.5398048	0	9071.00	49292

8:50 PM
4/3/2023

Figure 2: *FL Summary Statistics*

state=FL

Variable	Mean	Std Dev	Minimum	Maximum	N
InpatientBeds7DaySum	1801.67	2171.51	0	30695.00	29688
InpatientBedsUsed7DaySum	1351.57	1762.81	0	18622.00	29688
IBUR	0.7101467	0.2195058	0	2.2000000	29593
total_adult_patients_hospitalize	127.7361913	235.0515800	0	6020.00	28225

8:52 PM
4/3/2023

Figure 3: *NY Summary Statistics*

state=NY

Variable	Mean	Std Dev	Minimum	Maximum	N
InpatientBeds7DaySum	1911.35	2608.27	0	23744.00	23859
InpatientBedsUsed7DaySum	1439.86	2116.64	0	20505.00	23859
IBUR	0.7001228	0.1898866	0	1.2222222	23650
total_adult_patients_hospitalize	125.2097965	246.8289443	0	4689.00	22069

8:53 PM
4/3/2023

Figure 4: *TX Summary Statistics*

state=TX

Variable	Mean	Std Dev	Minimum	Maximum	N
InpatientBeds7DaySum	893.6039731	1490.11	0	16471.00	65238
InpatientBedsUsed7DaySum	673.0174898	1191.84	0	15349.00	65238
IBUR	0.5832428	0.3157934	0	2.8000000	63167
total_adult_patients_hospitalize	69.4513224	166.6235773	0	3328.00	61178

8:54 PM
4/3/2023

The summary statistics for available inpatient beds, inpatient beds used, inpatient bed usage rate (IBUR), and total number of adult patients hospitalized with COVID-19 are calculated as seen in Figures 1-4. The summary statistics include mean, standard deviation, minimum, and maximum values categorized by each relevant state. The average IBUR for CA, FL, and NY are relatively similar at around 70%. However, TX has the lowest IBUR among the largest populated states. FL and NY have the highest average number of hospitalized adults due to COVID at about 127 adults per hospital.

Figure 5: *Distributions of Available Inpatient Beds per State*

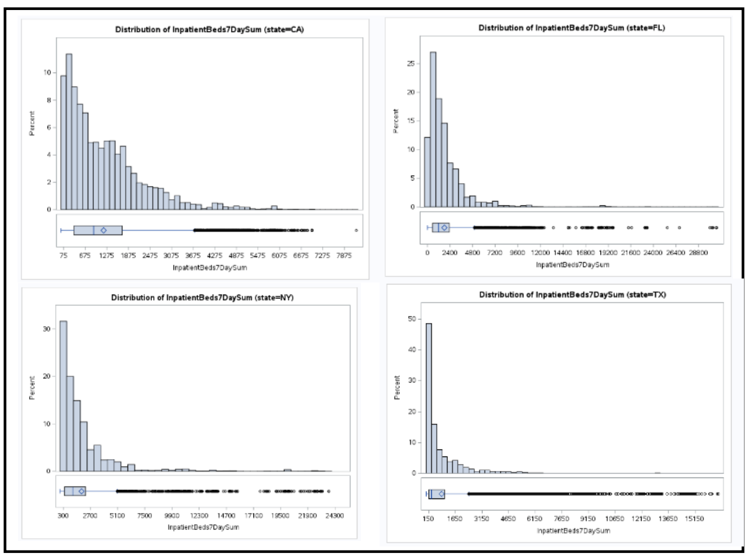


Figure 6: *Distributions of Inpatient Beds Used per State*

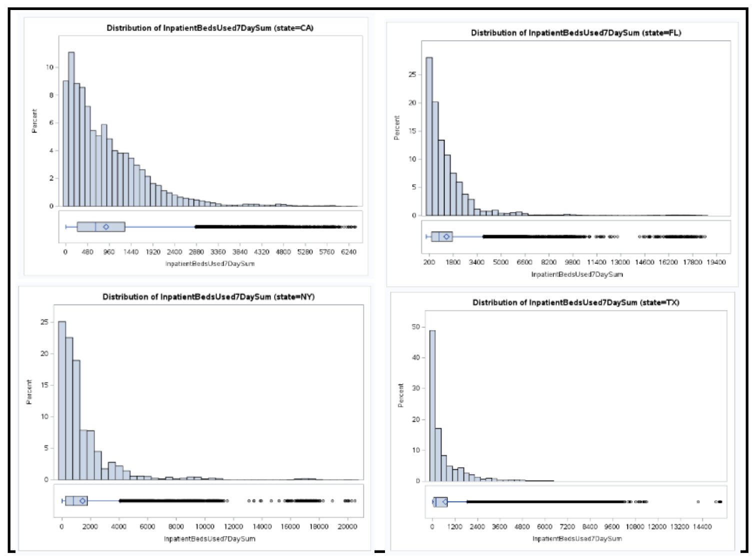


Figure 7: *Distribution of IBUR per State*

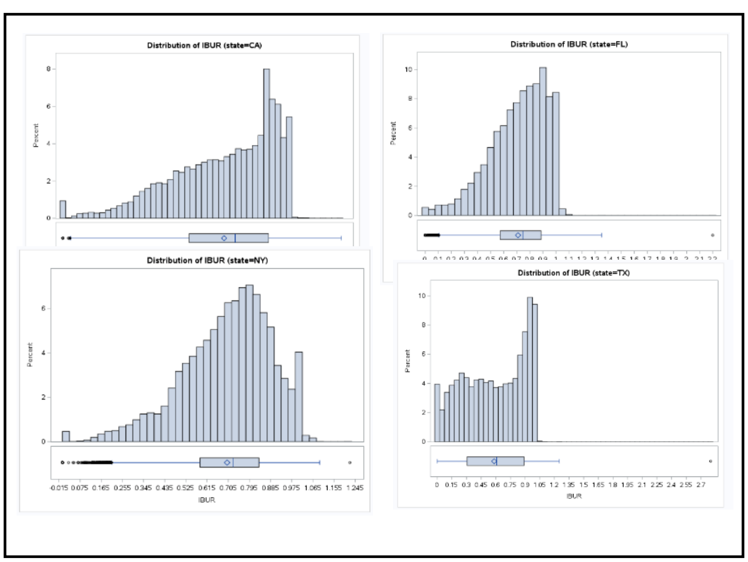
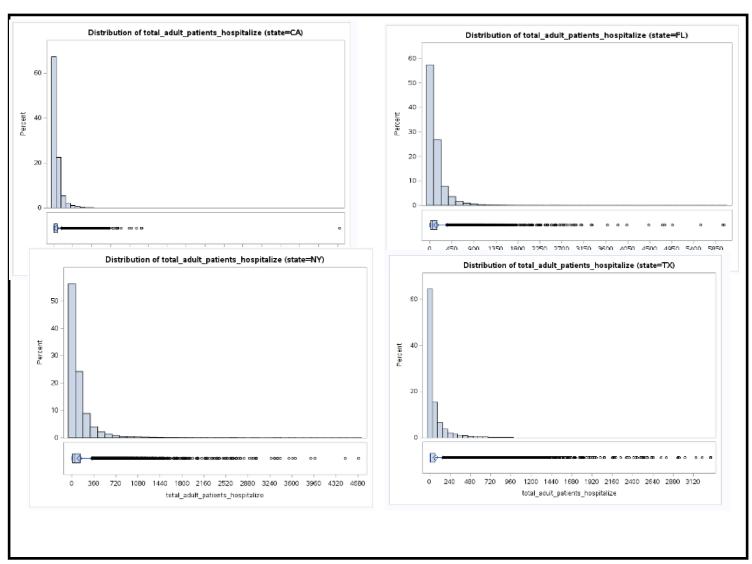


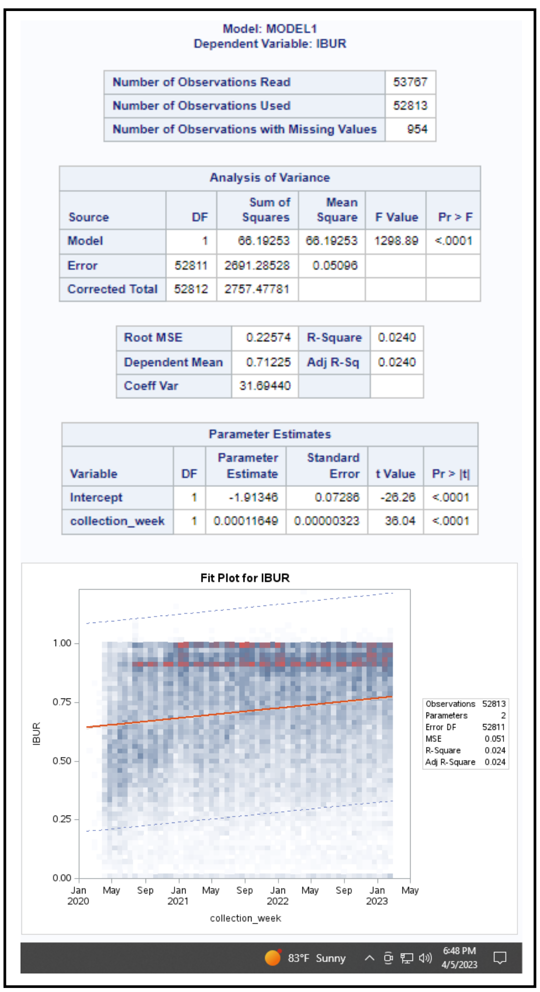
Figure 8: *Distribution of Hospitalized Adult COVID Patients per State*



Histograms and box plots for visualizing the distribution of each variable are displayed in Figures 5-8. The distribution for total inpatient beds available and inpatient beds used for all four states are skewed to the right as seen in Figure 5 and Figure 6. The box plots for each state and these two variables show the dataset contains multiple outliers. The histogram for the inpatient bed usage rate (IBUR) of each state shows the dataset is skewed to the left as most of the data is concentrated on the right side as seen in Figure 7. This signifies high IBUR were prevalent during the COVID-19 pandemic for all states. All histograms describing total adult hospitalized patients due COVID-19 are right tailed as seen in Figure 8.

Hospital Capacity Dataset Predictive Statistics

Figure 9: *CA's IBUR Time Regression Analysis*



Model: MODEL1
Dependent Variable: IBUR

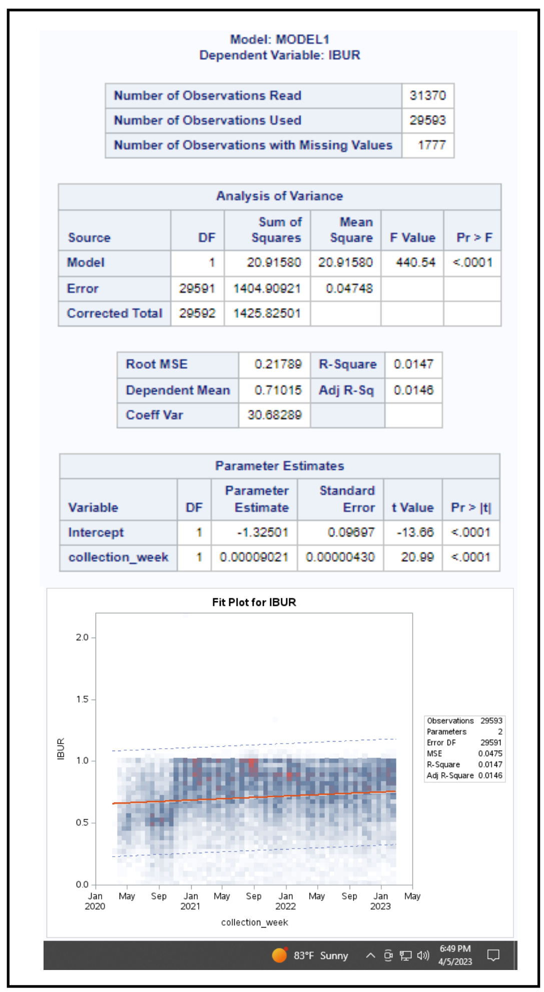
Number of Observations Read	53767
Number of Observations Used	52813
Number of Observations with Missing Values	954

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	66.19253	66.19253	1298.89	<.0001
Error	52811	2691.28528	0.05096		
Corrected Total	52812	2757.47781			

Root MSE	0.22574	R-Square	0.0240
Dependent Mean	0.71225	Adj R-Sq	0.0240
Coeff Var	31.69440		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-1.91346	0.07286	-26.26	<.0001
collection_week	1	0.00011649	0.00000323	36.04	<.0001

Figure 10: *FL's IBUR Time Regression Analysis*



Model: MODEL1
Dependent Variable: IBUR

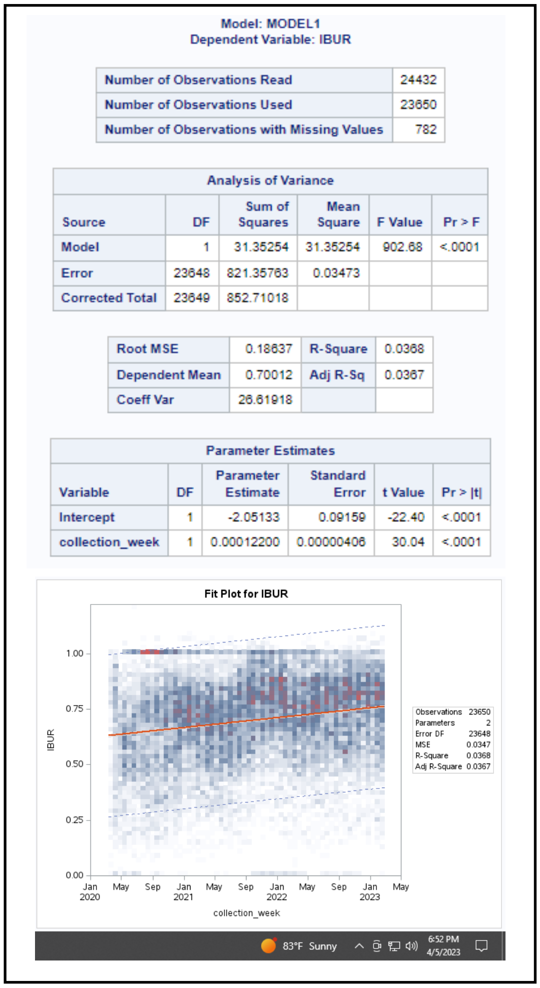
Number of Observations Read	31370
Number of Observations Used	29593
Number of Observations with Missing Values	1777

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	20.91580	20.91580	440.54	<.0001
Error	29591	1404.90921	0.04748		
Corrected Total	29592	1425.82501			

Root MSE	0.21789	R-Square	0.0147
Dependent Mean	0.71015	Adj R-Sq	0.0146
Coeff Var	30.68289		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-1.32501	0.09697	-13.66	<.0001
collection_week	1	0.00009021	0.00000430	20.99	<.0001

Figure 11: *NY's IBUR Time Regression Analysis*


Model: MODEL1
Dependent Variable: IBUR

Number of Observations Read	24432
Number of Observations Used	23650
Number of Observations with Missing Values	782

Analysis of Variance

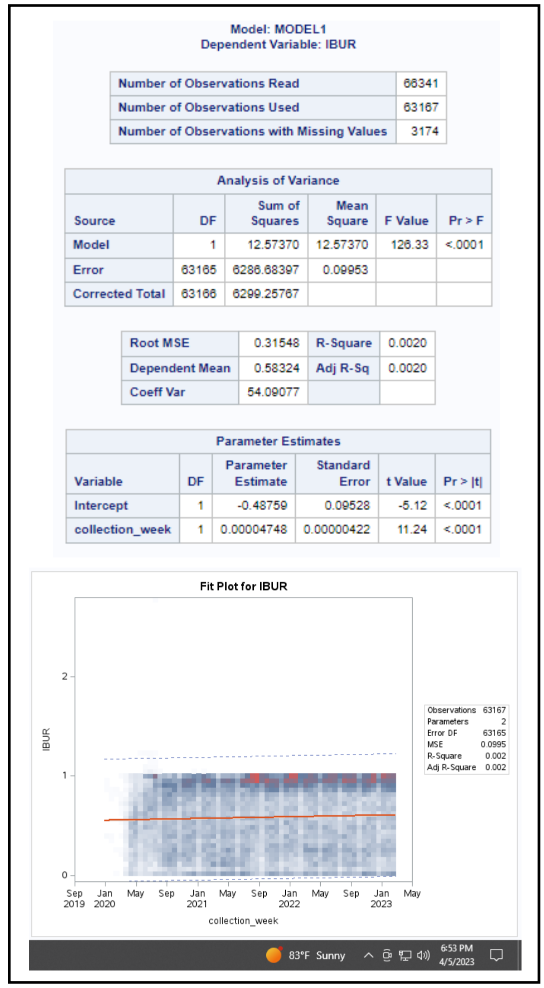
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	31.35254	31.35254	902.68	<.0001
Error	23648	821.35763	0.03473		
Corrected Total	23649	852.71018			

Root MSE	0.18637	R-Square	0.0368
Dependent Mean	0.70012	Adj R-Sq	0.0367
Coeff Var	26.61918		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-2.05133	0.09159	-22.40	<.0001
collection_week	1	0.00012200	0.00000406	30.04	<.0001

Figure 12: *TX's IBUR Time Regression Analysis*


Model: MODEL1
Dependent Variable: IBUR

Number of Observations Read	66341
Number of Observations Used	63167
Number of Observations with Missing Values	3174

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	12.57370	12.57370	126.33	<.0001
Error	63165	6286.68397	0.09953		
Corrected Total	63166	6299.25767			

Root MSE	0.31548	R-Square	0.0020
Dependent Mean	0.58324	Adj R-Sq	0.0020
Coeff Var	54.09077		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.48759	0.09528	-5.12	<.0001
collection_week	1	0.00004748	0.00000422	11.24	<.0001

Time regression analysis was performed for each state as seen in Figures 9-12. The data shows the IBUR steadily increased over the course of the pandemic spanning from January 2020 to the present. The correlation coefficients (R-square) are all extremely low. However, both the intercept and collection_week predictor are significant in each regression model as the p-values for all cases are less than 0.05. NY has the greatest rate of increasing IBUR with an increase of 0.000122 per week.

Figure 13: *CA's Hospitalized COVID Cases Regression Analysis*

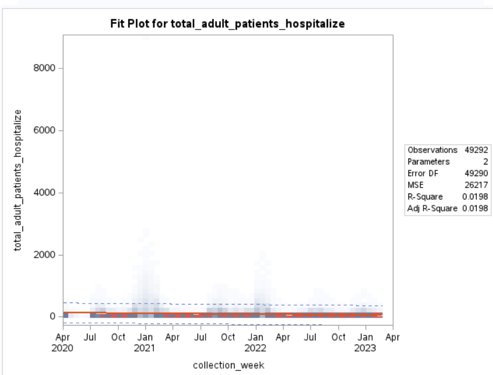
Model: MODEL1
Dependent Variable: total_adult_patients_hospitalize

Number of Observations Read	53767
Number of Observations Used	49292
Number of Observations with Missing Values	4475

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	26088350	26088350	995.11	<.0001
Error	49290	1292212643	26217		
Corrected Total	49291	1318300993			

Root MSE	161.91519	R-Square	0.0198
Dependent Mean	89.66240	Adj R-Sq	0.0198
Coeff Var	180.58315		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1915.72086	57.89132	33.09	<.0001
collection_week	1	-0.08088	0.00256	-31.55	<.0001



83°F Sunny 6:56 PM 4/5/2023

Figure 14: *FL's Hospitalized COVID Cases Time Time Regression Analysis*

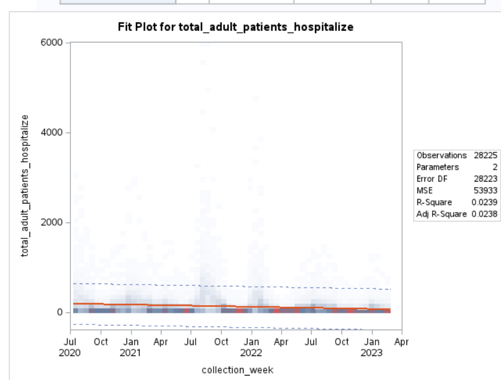
Model: MODEL1
Dependent Variable: total_adult_patients_hospitalize

Number of Observations Read	31370
Number of Observations Used	28225
Number of Observations with Missing Values	3145

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	37202469	37202469	689.79	<.0001
Error	28223	1522152228	53933		
Corrected Total	28224	1559354698			

Root MSE	232.23488	R-Square	0.0239
Dependent Mean	127.73619	Adj R-Sq	0.0238
Coeff Var	181.80821		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	3083.79926	112.56102	27.40	<.0001
collection_week	1	-0.13088	0.00498	-26.26	<.0001



83°F Sunny 6:57 PM 4/5/2023

Figure 15: *NY's Hospitalized COVID Cases Regression Analysis*

Model: MODEL1
Dependent Variable: total_adult_patients_hospitalize

Number of Observations Read	24432
Number of Observations Used	22069
Number of Observations with Missing Values	2363

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	818804	818804	13.45	0.0002
Error	22067	1343663675	60890		
Corrected Total	22068	1344482479			

Root MSE	246.75936	R-Square	0.0006
Dependent Mean	125.20980	Adj R-Sq	0.0006
Coeff Var	197.07672		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	629.26013	137.46408	4.58	<.0001
collection_week	1	-0.02231	0.00608	-3.67	0.0002

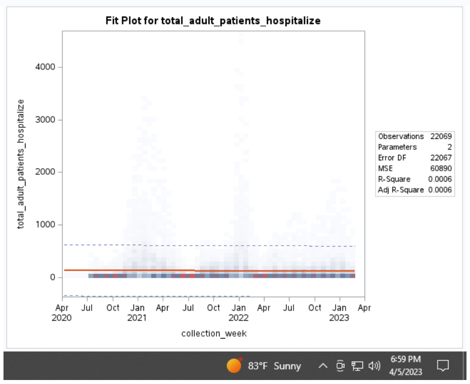


Figure 16: *TX's Hospitalized COVID Cases Time Time Regression Analysis*

Model: MODEL1
Dependent Variable: total_adult_patients_hospitalize

Number of Observations Read	66341
Number of Observations Used	61178
Number of Observations with Missing Values	5163

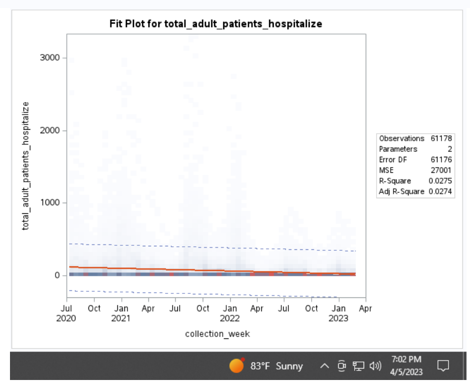
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	46647109	46647109	1727.58	<.0001
Error	61176	1651835422	27001		
Corrected Total	61177	1698482532			

Root MSE	164.32092	R-Square	0.0275
Dependent Mean	69.45132	Adj R-Sq	0.0274
Coeff Var	236.59869		

Parameter Estimates

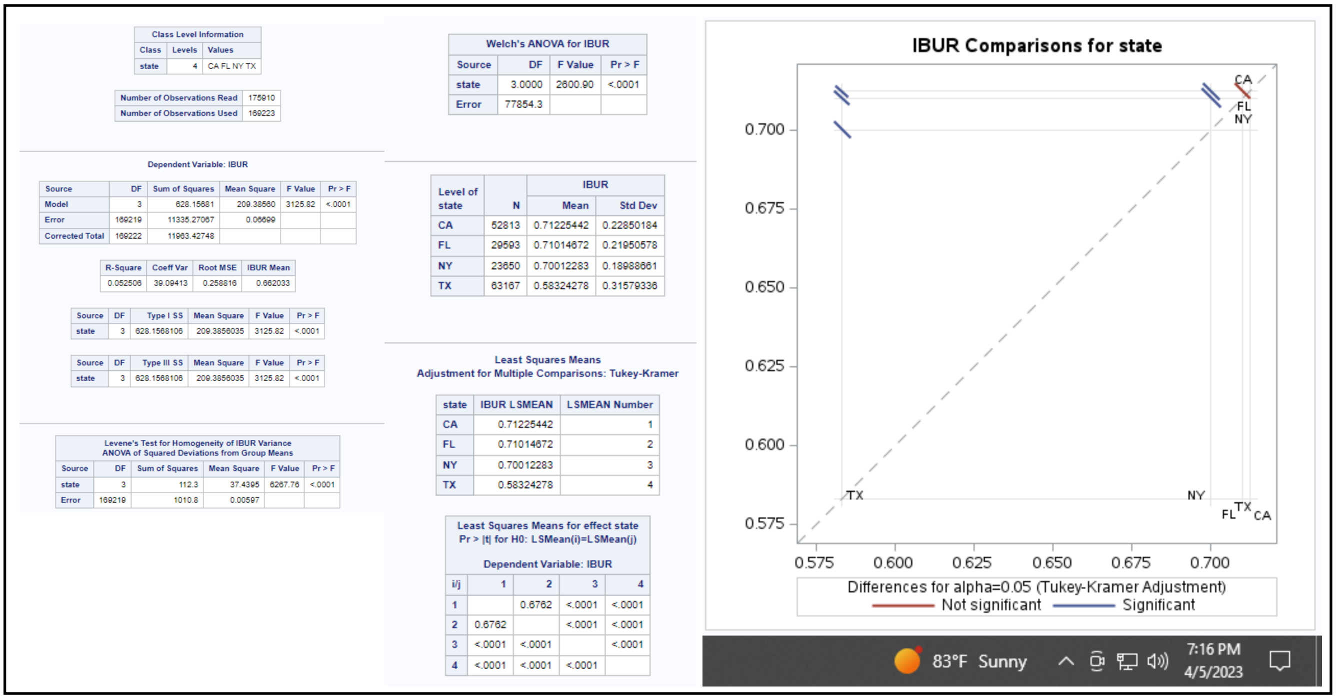
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2324.70900	54.26367	42.84	<.0001
collection_week	1	-0.09985	0.00240	-41.56	<.0001



Time regression analysis was performed for COVID-19 hospitalized adult cases for each state as seen in Figures 13-16. The data shows the total number of hospitalized adult patients due to COVID-19 steadily decreased over time for each state. The correlation coefficients (R-square) are all extremely low. However, both the intercept and collection_week predictor are significant in each regression model. NY has the slowest rate of decreasing number of hospitalized COVID-19 patients at -0.02231 patients per week.

Hospital Capacity Inferential Statistics

Figure 17: *One Way Analysis of Variance (ANOVA) for IBUR*



Class Level Information		
Class	Levels	Values
state	4	CA FL NY TX

Number of Observations Read	175910
Number of Observations Used	169223

Dependent Variable: IBUR

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	628.15681	209.38560	3125.82	<.0001
Error	169219	11335.27067	0.06699		
Corrected Total	169222	11963.42748			

R-Square	Coeff Var	Root MSE	IBUR Mean
0.052506	39.09413	0.258816	0.662033

Source	DF	Type I SS	Mean Square	F Value	Pr > F
state	3	628.1568106	209.3856035	3125.82	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
state	3	628.1568106	209.3856035	3125.82	<.0001

Levene's Test for Homogeneity of IBUR Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
state	3	112.3	37.4395	6267.76	<.0001
Error	169219	1010.8	0.00597		

Welch's ANOVA for IBUR			
Source	DF	F Value	Pr > F
state	3.0000	2600.90	<.0001
Error	77854.3		

Level of state	N	IBUR	
		Mean	Std Dev
CA	52813	0.71225442	0.22850184
FL	29593	0.71014672	0.21950578
NY	23650	0.70012283	0.18988661
TX	63167	0.58324278	0.31579336

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

state	IBUR LSMEAN	LSMEAN Number
CA	0.71225442	1
FL	0.71014672	2
NY	0.70012283	3
TX	0.58324278	4

Least Squares Means for effect state Pr > |t| for H0: LSMean(i)=LSMean(j) Dependent Variable: IBUR				
i/j	1	2	3	4
1		0.6762	<.0001	<.0001
2	0.6762		<.0001	<.0001
3	<.0001	<.0001		<.0001
4	<.0001	<.0001	<.0001	

A one-way analysis of variance test (ANOVA) was performed to test the first hypothesis that states inpatient bed usage rate (IBUR) is statistically different among the top 4 populated states. An alpha value of 0.05 was selected to represent the threshold of statistical significance. The Least Square Means for effect state table seen in Figure 17 supplies p-values where p-values less than 0.05 are significant. The graph IBUR Comparisons for State in Figure 17 also visualizes the significant and non-significant differences between states. The only non-significant difference appears between CA and FL. Both the visualization and least squared mean values support rejecting the null hypothesis as significant differences appear between some of the states. Therefore, there is a significant difference in the IBUR between some of the top populated states. CA and FL are seen to have the highest mean IBUR.

Figure 18: *One Way Analysis of Variance (ANOVA) for Total Hospitalized COVID Adult Patients*

Class Level Information		
Class	Levels	Values
state	4	CA FL NY TX

Number of Observations Read	175910
Number of Observations Used	160764

Dependent Variable: total_adult_patients_hospitalize

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	91380970	30460323	827.08	<.0001
Error	160760	5920620701	36829		
Corrected Total	160763	6012001670			

R-Square	Coeff Var	Root MSE	total_adult_patients_hospitalize Mean
0.015200	205.1720	191.9087	93.53550

Source	DF	Type I SS	Mean Square	F Value	Pr > F
state	3	91380969.55	30460323.18	827.08	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
state	3	91380969.55	30460323.18	827.08	<.0001

Levene's Test for Homogeneity of total_adult_patients_hospitalize Variance
ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
state	3	3.243E13	1.081E13	75.65	<.0001
Error	160760	2.297E16	1.429E11		

Welch's ANOVA for total_adult_patients_hospitalize

Source	DF	F Value	Pr > F
state	3.0000	689.36	<.0001
Error	62298.7		

Level of state	N	total_adult_patients_hospitalize Mean	total_adult_patients_hospitalize Std Dev
CA	49292	89.662400	163.539805
FL	28225	127.736191	235.051580
NY	22069	125.209797	246.828944
TX	61178	69.451322	166.623577

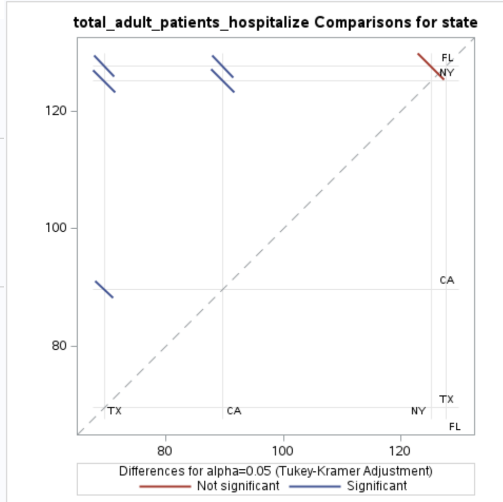
Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

state	total_adult_patients_hospitalize LSMEAN	LSMEAN Number
CA	89.662400	1
FL	127.736191	2
NY	125.209797	3
TX	69.451322	4

Least Squares Means for effect state
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: total_adult_patients_hospitalize

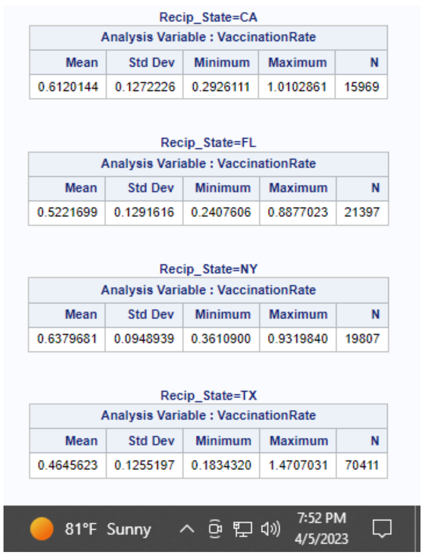
i/j	1	2	3	4
1		<.0001	<.0001	<.0001
2	<.0001		0.4588	<.0001
3	<.0001	0.4588		<.0001
4	<.0001	<.0001	<.0001	



A one-way analysis of variance test (ANOVA) was performed to test the second hypothesis that states there is statistical difference in the number of hospitalized adult COVID patients within top 4 populated states. An alpha value of 0.05 was selected to represent the threshold of statistical significance. The Least Square Means for effect state table seen in Figure 18 supplies p-values where p-values less than 0.05 are significant. The graph total_adult_patients_hosptialize Comparisons for State in Figure 18 also visualizes the significant and non-significant differences between states. The only non-significant difference appears between FL and NY. Both the visualization and least squared mean values support rejecting the null hypothesis as there is a significant difference between some states. Therefore, there is a significant difference in the average amount of hospitalized adult COVID patients between some of the top populated states. Although there is no significant difference between FL and NY, these states contain the most hospitalized adult COVID patients when compared to CA and TX.

Vaccinations Dataset Descriptive Statistics

Figure 19: *Vaccination Rate Summary Statistics for CA, FL, NY, and TX*



Recip_State=CA

Analysis Variable : VaccinationRate

Mean	Std Dev	Minimum	Maximum	N
0.6120144	0.1272226	0.2926111	1.0102861	15969

Recip_State=FL

Analysis Variable : VaccinationRate

Mean	Std Dev	Minimum	Maximum	N
0.5221699	0.1291616	0.2407606	0.8877023	21397

Recip_State=NY

Analysis Variable : VaccinationRate

Mean	Std Dev	Minimum	Maximum	N
0.6379681	0.0948939	0.3610900	0.9319840	19807

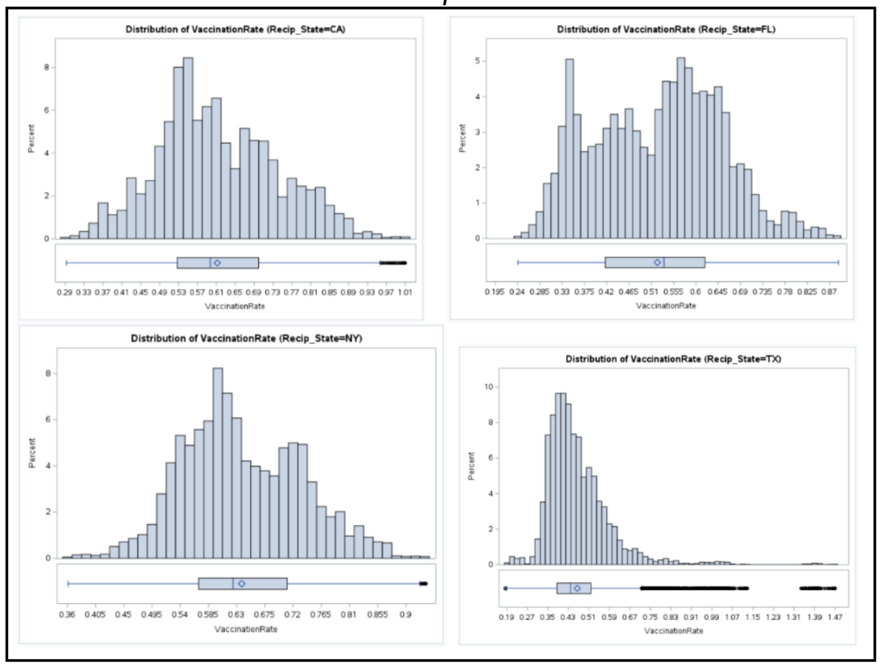
Recip_State=TX

Analysis Variable : VaccinationRate

Mean	Std Dev	Minimum	Maximum	N
0.4645623	0.1255197	0.1834320	1.4707031	70411

81°F Sunny 7:52 PM 4/5/2023

Figure 20: *Distribution of Vaccination Rate per State*



The summary statistics for the vaccination rates per state are seen Figures 19. The average vaccination rate for CA and NY seem to be much higher than TX upon initial review. The histograms and box plots for visualizing the vaccination rate distribution per state are displayed in Figure 20. The CA, FL, and NY seem to be somewhat bimodal while TX is skewed slightly to the right with many outliers.

Vaccinations Dataset Predictive Statistics

Figure 21: *CA's Vaccination Rate Time Regression Analysis*

Model: MODEL1
Dependent Variable: VaccinationRate

Number of Observations Read	15969
Number of Observations Used	15969

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	21.34813	21.34813	1437.63	<.0001
Error	15967	237.10324	0.01485		
Corrected Total	15968	258.45137			

Root MSE	0.12186	R-Square	0.0826
Dependent Mean	0.61201	Adj R-Sq	0.0825
Coeff Var	19.91111		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-6.04062	0.17546	-34.43	<.0001
Date	1	0.00029294	0.00000773	37.92	<.0001

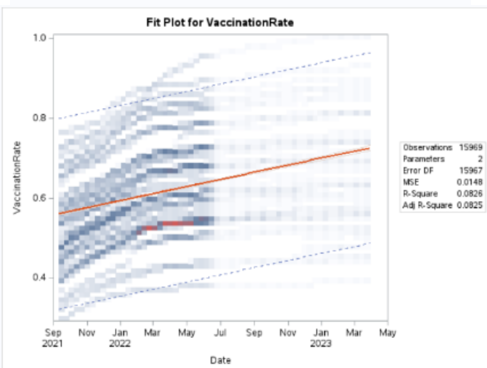


Figure 22: *FL's Hospitalized COVID Cases Time Time Regression Analysis*

Model: MODEL1
Dependent Variable: VaccinationRate

Number of Observations Read	21397
Number of Observations Used	21397

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	14.30032	14.30032	892.93	<.0001
Error	21395	342.64289	0.01602		
Corrected Total	21396	356.94321			

Root MSE	0.12655	R-Square	0.0401
Dependent Mean	0.52217	Adj R-Sq	0.0400
Coeff Var	24.23555		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-4.18179	0.15742	-26.56	<.0001
Date	1	0.00020713	0.00000693	29.88	<.0001

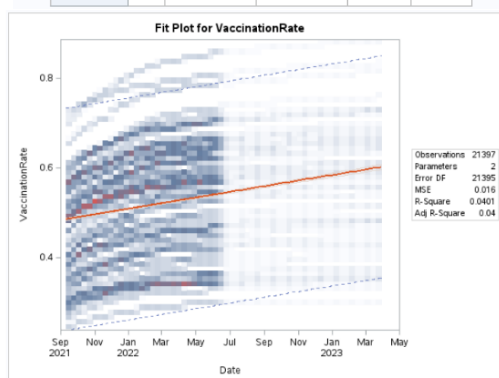
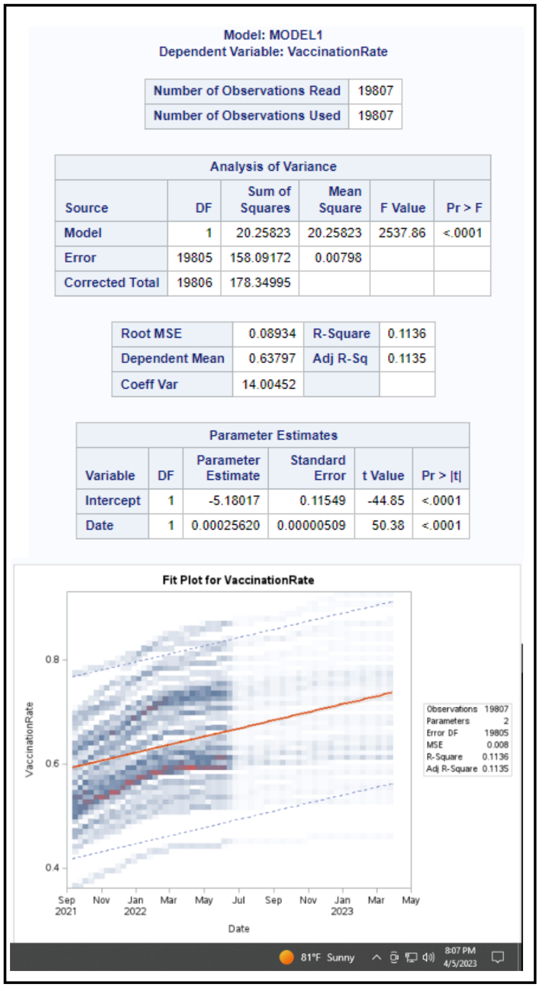


Figure 23: *NY's Vaccination Rate Time Regression Analysis*



Model: MODEL1
Dependent Variable: VaccinationRate

Number of Observations Read	19807
Number of Observations Used	19807

Analysis of Variance

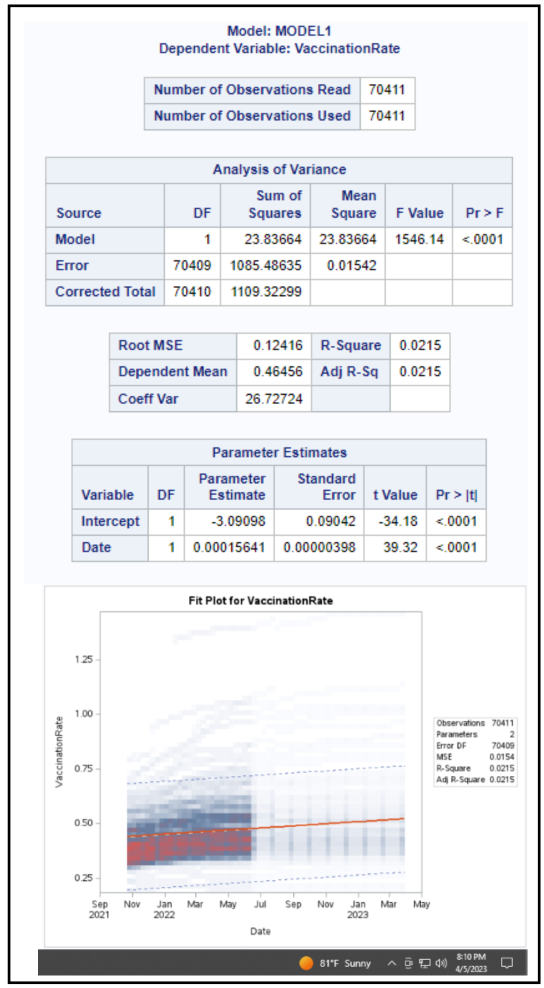
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	20.25823	20.25823	2537.86	<.0001
Error	19805	158.09172	0.00798		
Corrected Total	19806	178.34995			

Root MSE	0.08934	R-Square	0.1136
Dependent Mean	0.63797	Adj R-Sq	0.1135
Coeff Var	14.00452		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-5.18017	0.11549	-44.85	<.0001
Date	1	0.00025620	0.00000509	50.38	<.0001

Figure 24: *TX's Hospitalized COVID Cases Time Time Regression Analysis*



Model: MODEL1
Dependent Variable: VaccinationRate

Number of Observations Read	70411
Number of Observations Used	70411

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	23.83664	23.83664	1546.14	<.0001
Error	70409	1085.48635	0.01542		
Corrected Total	70410	1109.32299			

Root MSE	0.12416	R-Square	0.0215
Dependent Mean	0.46456	Adj R-Sq	0.0215
Coeff Var	26.72724		

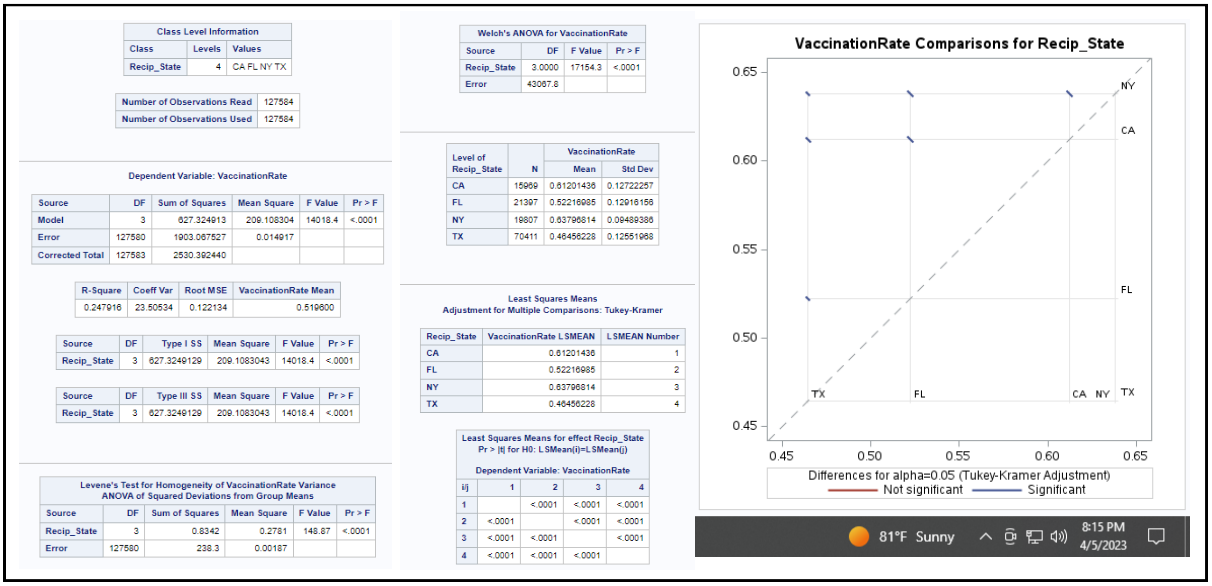
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-3.09098	0.09042	-34.18	<.0001
Date	1	0.00015641	0.00000398	39.32	<.0001

Time regression analysis was performed for vaccination rate as seen in Figures 21-24. The data shows the vaccination rate steadily increased over time for each state. The correlation coefficients (R-square) are all extremely low. However, both the intercept and collection_week predictor are significant in each regression model as the p-values are less than 0.05. TX has the lowest rate of vaccination improvement at 0.000156 per week.

Vaccination Dataset Inferential Statistics

Figure 25: *One Way Analysis of Variance (ANOVA) for Total Hospitalized COVID Adult Patients*



Class Level Information		
Class	Levels	Values
Recip_State	4	CA FL NY TX

Number of Observations Read	127584
Number of Observations Used	127584

Dependent Variable: VaccinationRate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	627.324913	209.108304	14018.4	<.0001
Error	127580	1903.067527	0.014917		
Corrected Total	127583	2530.392440			

R-Square	Coeff Var	Root MSE	VaccinationRate Mean
0.247916	23.50534	0.122134	0.519600

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Recip_State	3	627.3249129	209.1083043	14018.4	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Recip_State	3	627.3249129	209.1083043	14018.4	<.0001

Levene's Test for Homogeneity of VaccinationRate Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Recip_State	3	0.8342	0.2781	148.87	<.0001
Error	127580	238.3	0.00187		

Welch's ANOVA for VaccinationRate			
Source	DF	F Value	Pr > F
Recip_State	3.0000	17154.3	<.0001
Error	43067.8		

Level of Recip_State	N	VaccinationRate	
		Mean	Std Dev
CA	15969	0.61201436	0.12722257
FL	21397	0.52216985	0.12916156
NY	19807	0.63796814	0.09489386
TX	70411	0.46456228	0.12551968

Least Squares Means
Adjustment for Multiple Comparisons: Tukey-Kramer

Recip_State	VaccinationRate LSMEAN	LSMEAN Number
CA	0.61201436	1
FL	0.52216985	2
NY	0.63796814	3
TX	0.46456228	4

Least Squares Means for effect Recip_State Pr > |t| for H0: LSMean(i)=LSMean(j) Dependent Variable: VaccinationRate				
i/j	1	2	3	4
1		<.0001	<.0001	<.0001
2	<.0001		<.0001	<.0001
3	<.0001	<.0001		<.0001
4	<.0001	<.0001	<.0001	

A one-way analysis of variance test (ANOVA) was performed to test the third hypothesis that states there is statistical difference in vaccination rate among the top 4 populated states. The Least Square Means for effect state table seen in Figure 25 supplies p-values where p-values less than 0.05 are significant. The differences between all states are significant. Both the visualization and least squared mean values support rejecting the null hypothesis. Therefore, there is a significant difference in the vaccination rate between the top populated states. NY is seen to have the highest mean vaccination rate of 63.8% while TX has the lowest vaccination rate of 46.5%.

Conclusions

The purpose of this research project is to evaluate resource management methods associated with COVID-19. The US struggled to manage overflowing hospital capacity and vaccine distribution during the height of the pandemic. This research project lists three hypotheses associated with understanding how the US managed their resources. The first hypothesis addresses if the inpatient bed usage rate (IBUR) between the top four populated states is significantly different. The ANOVA study found that there is a significant difference in the IBUR between the top populated states where CA and FL are seen to have the highest average IBUR. IBUR increased over the entire existence of the pandemic for all states according to time regression analysis.

The second hypothesis states the average number of hospitalized adult COVID patients within top 4 populated states is statistically different. The ANOVA study found that this metric to be statistically different between all states except for the non-significant difference between FL and NY. FL and NY had a higher average amount of hospitalized COVID patients when compared to the other states. The time regression analysis shows that the average number of hospitalized COVID cases decreased over time for each state.

The third and final hypothesis states that the average vaccination rate within the top four populated states is statistically different. The ANOVA study found that there is a statistical difference between all four states. NY had the highest vaccination rate while TX had the lowest vaccination rate.

Recommendations

Due to the high IBUR for both CA and FL, the United States may want to investigate other factors that contribute to these high rates. At the very least, additional resources should be supplied to CA and FL to mitigate their respective IBURs. For more granular analyses, further studies can be conducted on the impact of seasonal changes on IBUR. Additional research could be performed to analyze factors attributed to the high amount of hospitalized adult COVID patients within FL and NY. This may include studies on population density per state. The United States should promote vaccination programs within TX to improve their vaccination rate. They should look to factors that make NY successful in this area.

References

Alkarkhi, A. (2021). 1 - Introduction to statistical hypothesis testing*, Applications of Hypothesis Testing for Environmental Science*, Elsevier, Pages 1-12, ISBN 9780128243015, https://doi.org/10.1016/B978-0-12-824301-5.00006-X.

Brüggemann, S., Chan, T., Wardi, G., Mandel, J., Fontanesi, J., & Bitmead, R. R. (2021). Decision support tool for hospital resource allocation during the COVID-19 pandemic. *Informatics in Medicine Unlocked,* 24. https://doi.org/10.1016/j.imu.2021.100618

Centers for Disease Control and Prevention (2023) *COVID-19 Vaccinations in the United States, County* [Dataset]. https://healthdata.gov/dataset/COVID-19-Vaccinations-in-the-United-States-County/ipdn-uaih

Darity, W. (2008). Time Series Regression. *International Encyclopedia of the Social Sciences*, 2nd ed., Vol. 8, p. 371-372. Macmillan Reference USA.

Jensen, A. (2020). *COVID-19: Overview, Economic Implications and Federal Response.* SNOVA.

Scheff, S. (2016) Chapter 6 - One-Way Analysis of Variance, *Fundamental Statistical Principles for the Neurobiologist*, Academic Press, Pages 97-133, ISBN 9780128047538, https://doi.org/10.1016/B978-0-12-804753-8.00006-3.

Shukla, S., Fressin, F., Un, M., Coetzer, H., & Chaguturu, S. K. (2022). Optimizing vaccine distribution via mobile clinics: a case study on COVID-19 vaccine distribution to long-term care facilities. *Vaccine, 40(5),* 734–741. https://doi.org/10.1016/j.vaccine.2021.12.049

Singh, S., Ambooken, G. C., Setlur, R., Paul, S. K., Kanitkar, M., Singh Bhatia, S., & Singh Kanwar, R. (2021). Challenges faced in establishing a dedicated 250 bed COVID-19 intensive care unit in a temporary structure. *Trends in Anaesthesia and Critical Care,* 36, 9–16. https://doi.org/10.1016/j.tacc.2020.10.006

US Department of Health and Human Services (2023) *COVID-19 Reported Patient Impact and Hospital Capacity by Facility* [Dataset]. https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/anag-cw7u