ALY 6060 – Decision Support and Business Intelligence

Instructor: Vera Yanovsky

Northeastern University

Final Project Report

Created By:

Karan Daiya
Novruz Guliyev
Sandeep Karapa Srivatsav
Sweta Mankala

Date: 07/02/2021

INTRODUCTION

The COVID-19 dataset we picked is accessible on the Massachusetts Department of Public Health's official website, mass.gov. The dataset comprises 31 spreadsheets in total, and we picked the second provider, "Academic Medical Center," for our research. This is raw data that includes covid data on a variety of public health topics. Other matrices, such as average, median cases, and fatalities, are also used, as well as other periods, such as months and weeks. The file is in the xlsx format. It contains several tabs for different aspects including age, data factorization by city, county, and so on. Furthermore, this study will focus on various aspects of public health and will aid the University Systems' Health Science Sector in responding to a range of questions and issues. UT Physicians collaborates with the Harris Health System (HHS) and the Memorial Hermann Healthcare System (MHHS). They also provide an Electronic Health Record (EHR) system that meets all of UT Physicians' demands while also enhancing the patient and provider experience. We might plan the different approaches that could be improved and applied, defining the profits to be made because UTHSC-T intends to educate on professional training for students in terms of research and patient care. As a team of Data Analysts, we work to deliver insights for our provider the Academic Medical Center which is "The Health Science Center". We would like to address metrics to provide resources to medical schools, hospitals, and research facilities present in the Health Science Center with most clinical operations in the medical school practice plan. With the health science center allocating revenue for the resources, we aim to break down some of the key metrics we have identified for the number of cases, hospitalizations, resource utilization for the patients. This breakdown will help in creating a solution with a system of Electronic Health Records for patients in organizations to provide resources and determine budget and revenue for better patient and hospitalization experience.

BUSINESS USE CASE AND OBJECTIVES

1. What is the patient breakdown in terms of probable cases, hospitalizations, and deaths by age groups, race & ethnicity, and sex? This question will aid UT Physicians in focusing their efforts on patients who are more critical at this time.

2. What to expect from hospitalizations and how quickly people are getting recovered and how many deaths occurred after hospitalizations?

3. What is the number of patients to be accommodated to provide the resources for hospitalization? Also, does the rise in the number of cases necessarily indicate the number of resources to be provided to the patients?

4. What kind of hospitalization or treatment is given to patients? This will help the medical center to be better prepared for any rise of cases in the future.

How does the visualization aid the intended audience in decision support?

1. The very first dashboard represents quick visualization of numbers that elaborates total active cases, total deaths, current patients in hospital, and most affected race till now. Furthermore, the dashboard represents the last 2 weeks' KPIs which states the most affected age group, most hospitalized age group, and most affected sex group. The most interactive part is that a user can hover over the blue digits with "**" for more detailed insights. Last, there are two area charts which state the "Positive new cases vs Estimated active cases" and "Daily Hospitalizations."

Total Active Cases	* Total Deaths *	Current Patients in Hospital	* Most Affected Race/Ethnicity *
663,666	17,938	95	White, non-Hispanic 271,164

*(Note: Hover over the blue digits in boxes with " ** " for further insights)*

Last 2 Weeks KPIs :

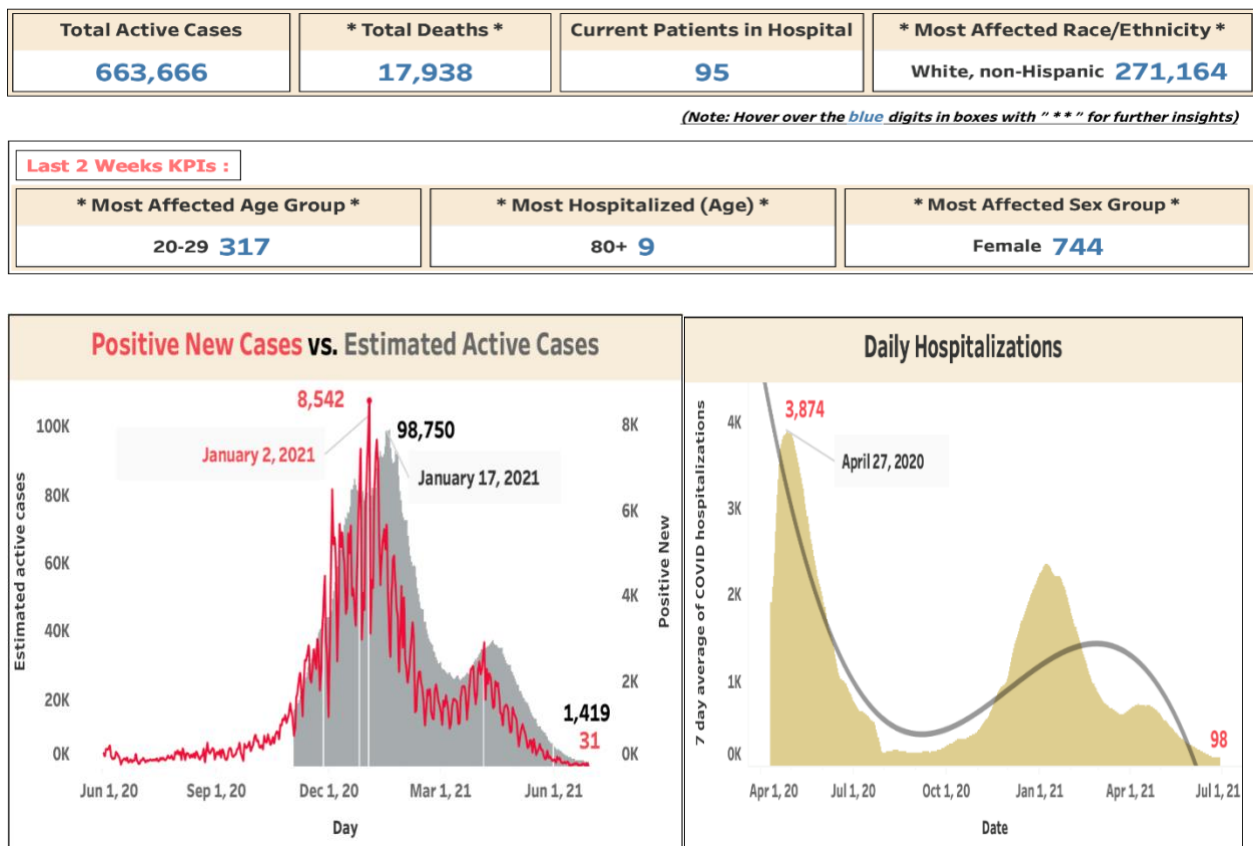| * Most Affected Age Group *	* Most Hospitalized (Age) *	* Most Affected Sex Group *
20-29 317	80+ 9	Female 744

Fig 1. Dashboarding of Patient Breakdown

2. Second dashboard represents the relationships between hospitalization and death. We have tried to find what is the actual relationship between death rate and hospitalization. First,

top left visualization aims to find what would be death rate in a specific number of hospitalization cases. Relationship is not a linear which means hospitalization and death rate does not increase linearly. Relationship is not exponential too. We have approximated the relationship with 3rd degree polynomial curve. Second visualization represents number of death cases and hospitalization over months. We see that in summer times people are less likely to death and hospitalization. Third visualization demonstrates the difference between races for their death rate and hospitalization.

The following graphs will demonstrate the level of hospitalization. Level of Death in Hospitalization over months, Density plot shows density of hospitalization, 3-degree polynomial line shows death number for those days.
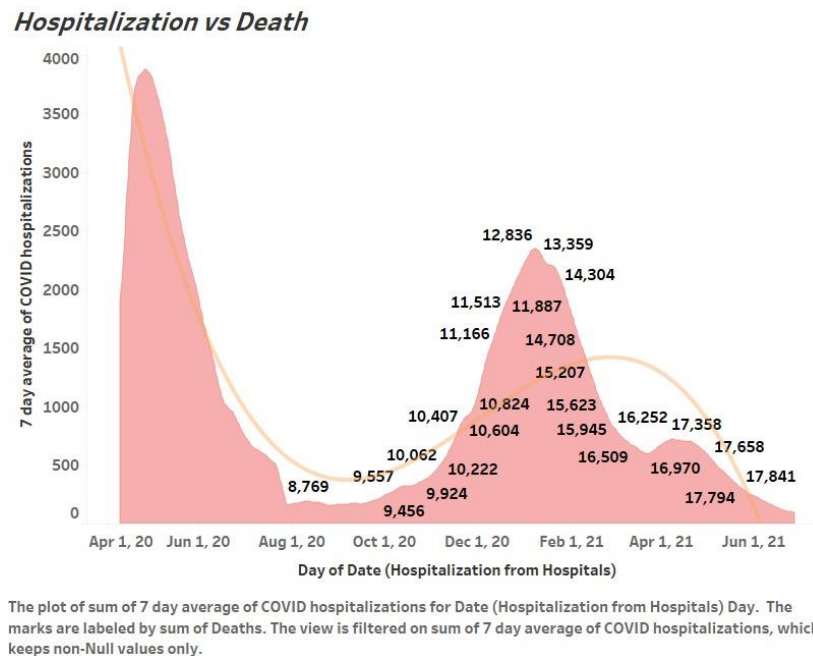

The plot of sum of 7 day average of COVID hospitalizations for Date (Hospitalization from Hospitals) Day. The marks are labeled by sum of Deaths. The view is filtered on sum of 7 day average of COVID hospitalizations, which keeps non-Null values only.

Fig 2. Hospitalization vs. Deaths

Another interesting pattern regarding hospitalization is to check which months have higher hospitalization.
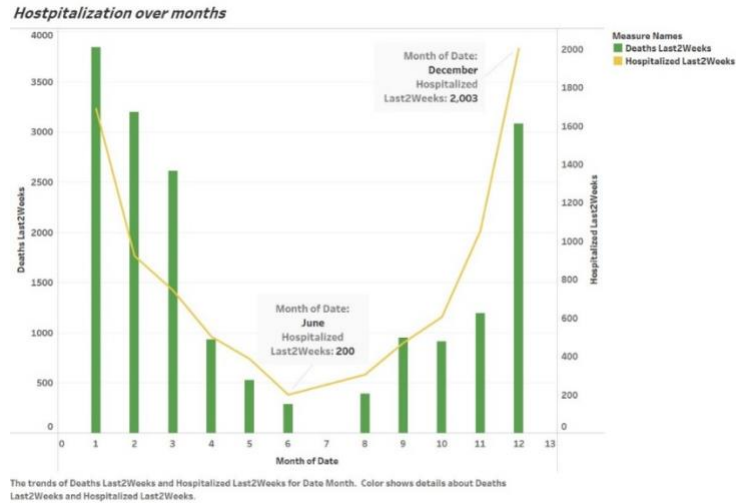
Hostpitalization over months

Fig 3. Hospitalization over Months

We see that summer times have lower hospitalization and a low death rate. This may be due to different reasons. Maybe people spend more time outside. The virus is less likely to spread outside. In June 200 hospitalization was recorded but in December this number is 2003.
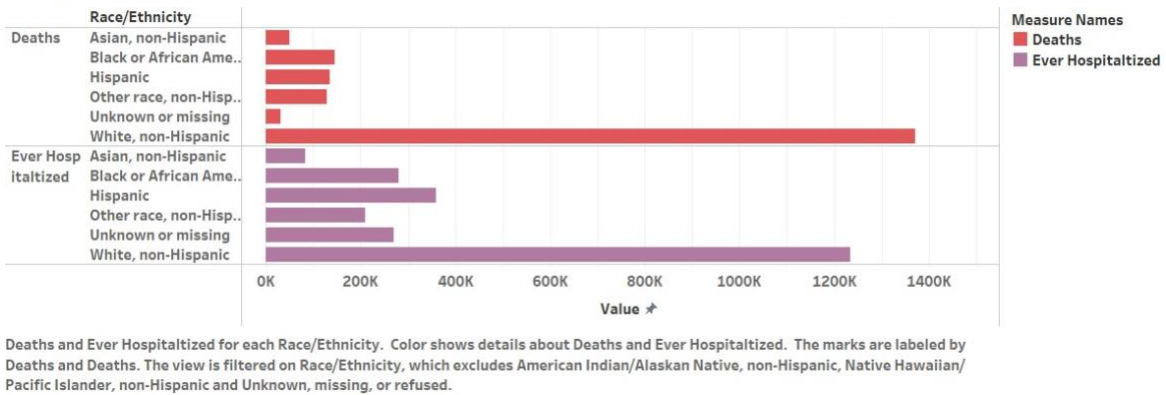
Hospitalization and Death over Race

Deaths and Ever Hospitaltized for each Race/Ethnicity. Color shows details about Deaths and Ever Hospitaltized. The marks are labeled by Deaths and Deaths. The view is filtered on Race/Ethnicity, which excludes American Indian/Alaskan Native, non-Hispanic, Native Hawaiian/ Pacific Islander, non-Hispanic and Unknown, missing, or refused.

Fig 4. Hospitalization and Death among different groups

The above graph shows death and hospitalization over races. White and non-Hispanic races have a higher number of hospitalizations compare with other races. We cannot directly allege that White race people are more likely to have coronavirus. Maybe, other races prefer treatment at home.
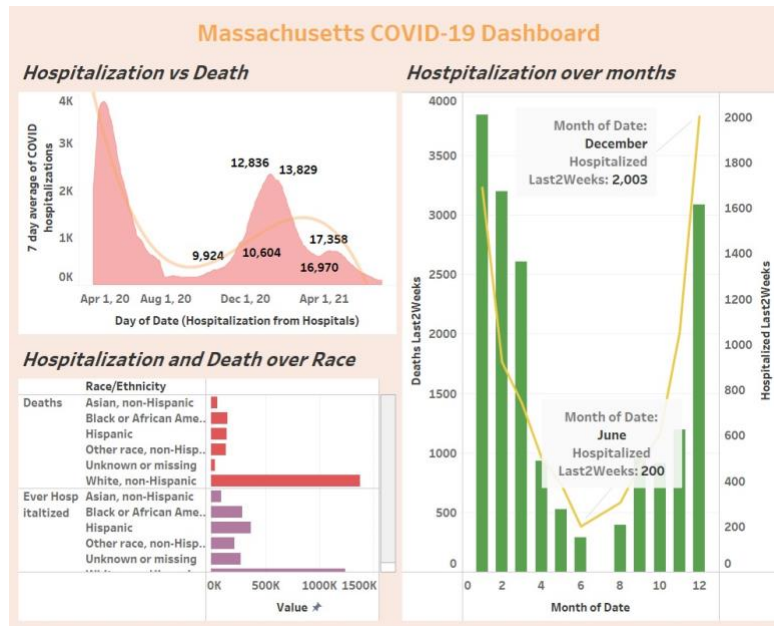
Fig 5. Dashboarding of Hospitalization vs. Deaths

3. The third dashboard provides information central to the number of cases and tests that are conducted with the analysis of the hospitalizations and their trend across the time of the pandemic.

The following graph shows the number of tests conducted indicated as the *Molecular Total* and the number of positive cases that were found from the tests conducted indicated as *Positive Total*. The graph shows a significant rise in the total number of tests conducted from after July 2020 onwards that correlates with the positive cases.
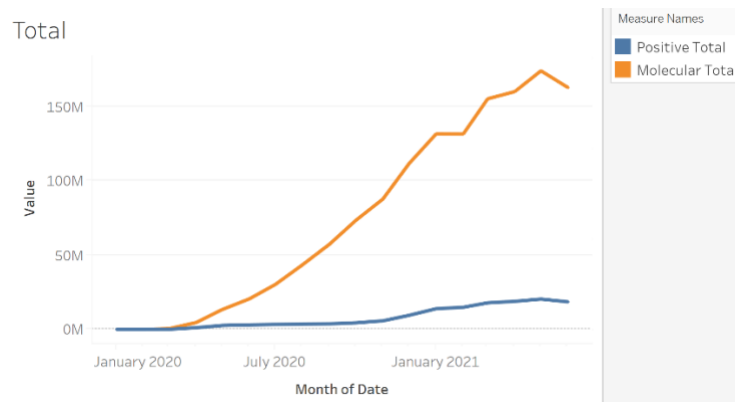
Fig 6. Total number of tests and positive cases

The following graph shows the number of positive cases encountered daily. The number of tests conducted was more in the months between December 2020 to January 2021.
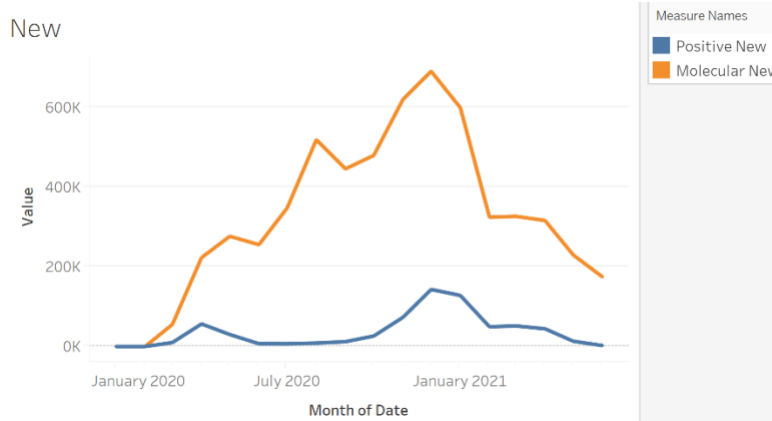

Fig 7. Daily tests and confirmed positive cases

The following graphs provide information on the number of new positive cases in patients with the number of patients hospitalized. There were a lot of hospitalizations witnessed in May 2020 but not many positive cases and a significant rise from December to January 2021 but not as many hospitalizations as witnessed earlier. The number of new COVID-19 hospitalizations shows a rise with the number of positive cases encountered for December and January 2021. The scatterplot between the new COVID patients in the hospital vs. the number of positive cases and 7-day confirmed cases indicate the highest number of cases recorded in January 2021 in Massachusetts.
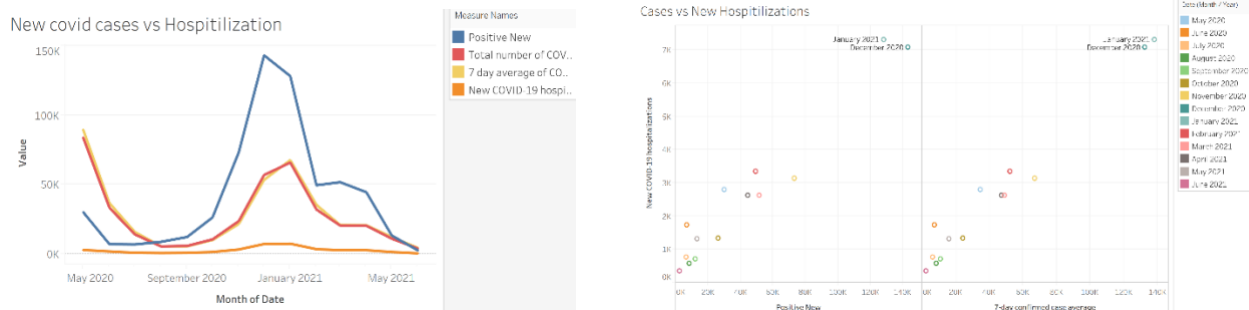

Fig 8. New COVID cases vs. hospitalizations

The following graph shows overall information on the hospitalizations on a 7-day average. The number of hospitalizations was more in April and May 2020 with not many confirmed cases. As testing increased over the period, there is a significant rise in confirmed cases between December 2020 to January 2021. The new COVID-19 hospitalizations correlate with the confirmed cases and the total number of patients.
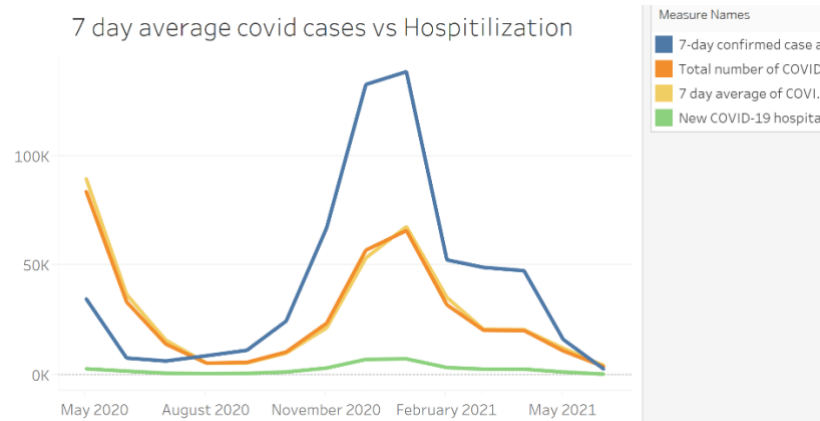

Fig 9. Analysis of average COVID cases per week

The following graph shows the total number of COVID patients in the hospital daily concerning the positive cases and the 7-day confirmed cases. The number of cases was high during the months December 2020 and January 2021 indicating a lagging factor for both the cases which proves enough information to the previous graph.
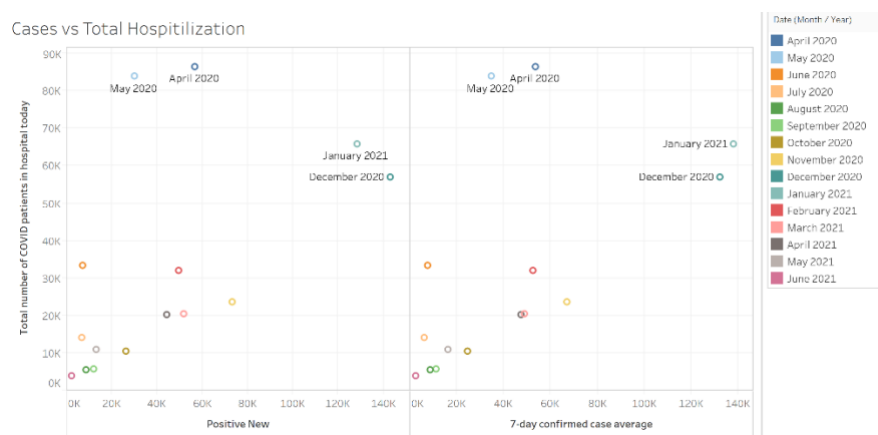

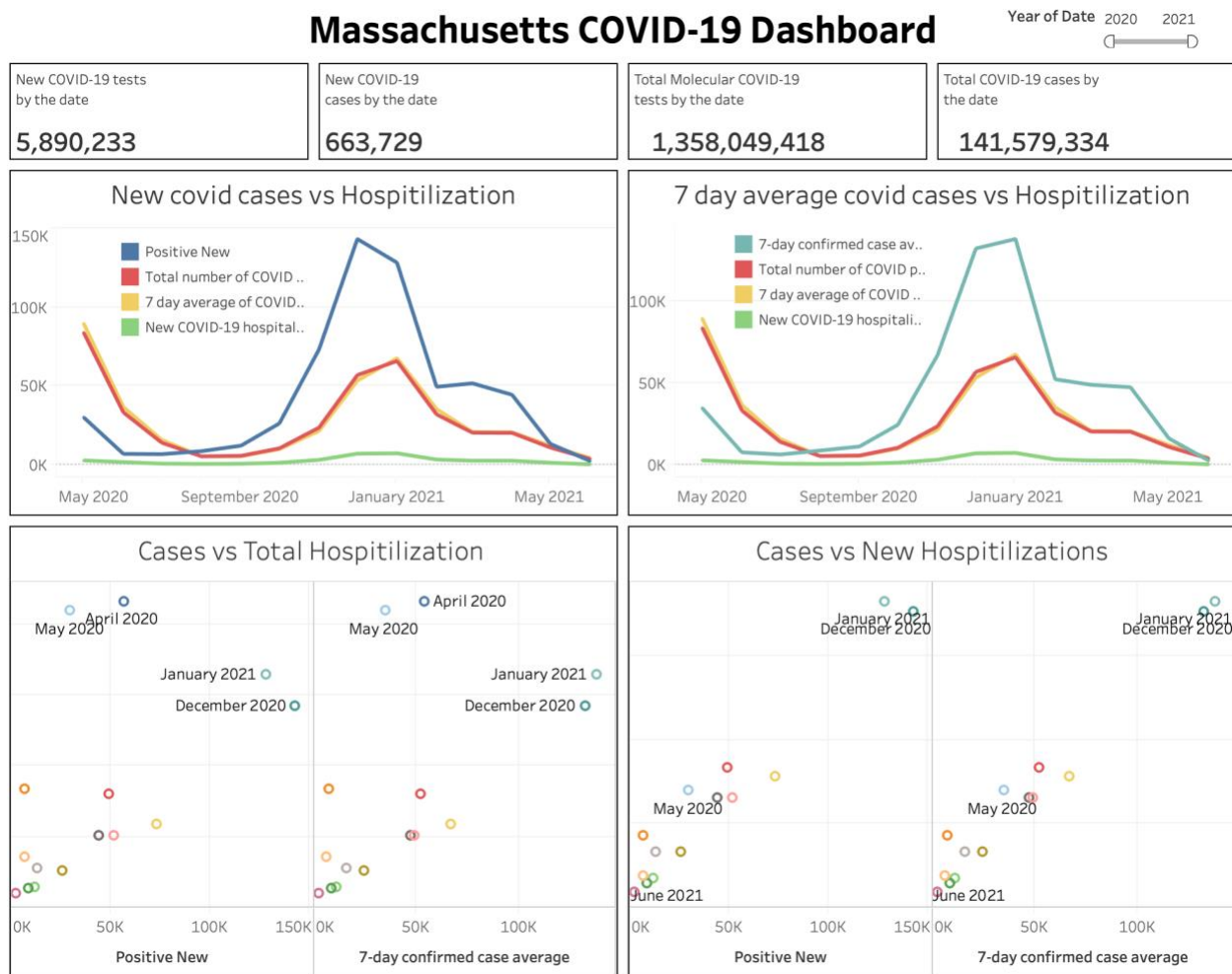Fig 10. Cases vs. Total hospitalizations

Fig 11. Dashboarding of Cases vs. Hospitalization

4. The fourth dashboard provides information on the kind of hospitalizations that can be provided for the patients being affected by COVID and how they can be prepared for any kind of rise in cases. Hospitals and health care departments that oversee the resources can utilize this dashboard to see the trend in the kind of beds that were used by patients.

The following graph shows the trend of the number of ICU beds used across a period using available and occupied ICUs with the total number of COVID patients and the 7-day confirmed cases. With the number of COVID cases being at the peak in April 2020, we can see the number of occupied ICU beds also being used the most in the following months between May 2020 to

November 2020. Also, the new covid cases rise in the months December 2020 to January 2021, there is a peak in the number of occupied ICU beds. The availability of ICU beds shrank from December 2020. The scatterplot confirms the statistics with December 2020 and January 2021 having the greatest number of cases in patients occupying the ICU beds.
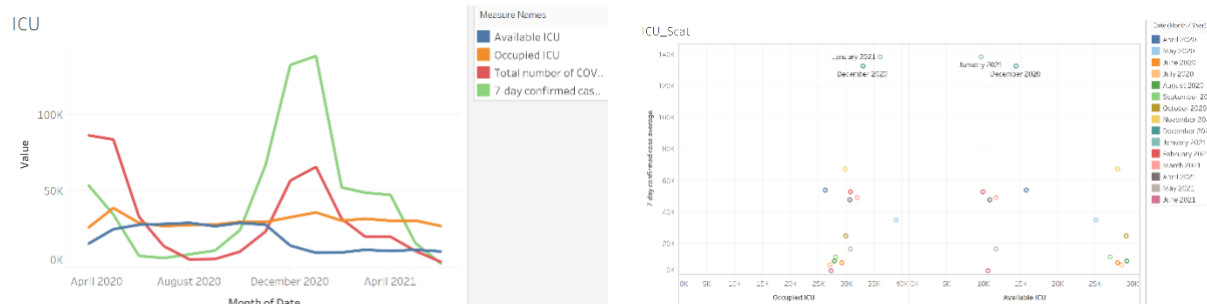

Fig 12. Deep dive analysis of trend in ICU usage

The following is the trend for the number of surgical beds used which shows the use beds are the most since May 2020. The trend had a dip in February and again rose. The dip in the use of surgical beds correlates with the availability of ICU beds. The scatterplot represents the same information with the number of surgical beds being used the most in December 2020 and January 2021 along with many COVID patients.
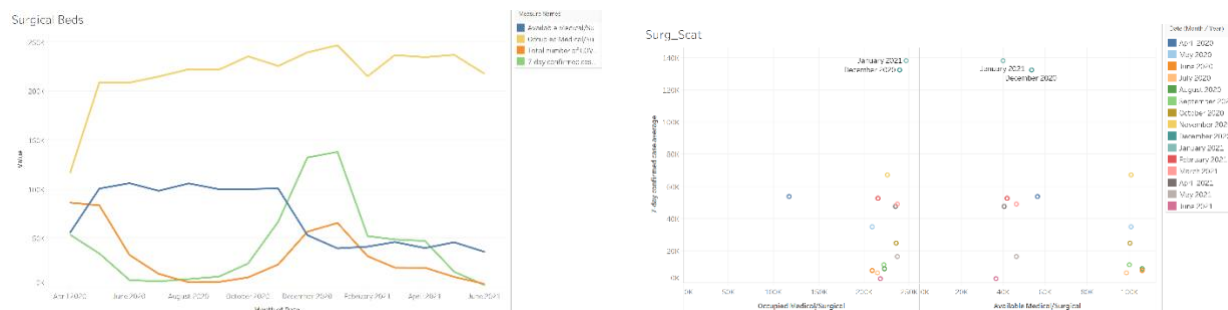

Fig 13. Deep dive analysis of trend of Surgical bed usage

The following trend graph shows the number of alternate beds highly occupied in February 2021 that correlates to patients being discharged from ICU or surgical beds to alternate medical beds.

The scatterplot shows the availability of beds being the most in February 2021 and the highest usage in January 2021.
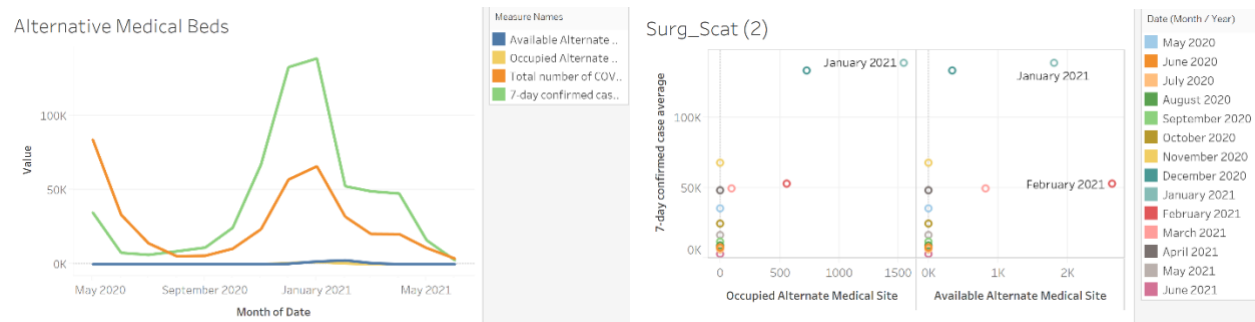

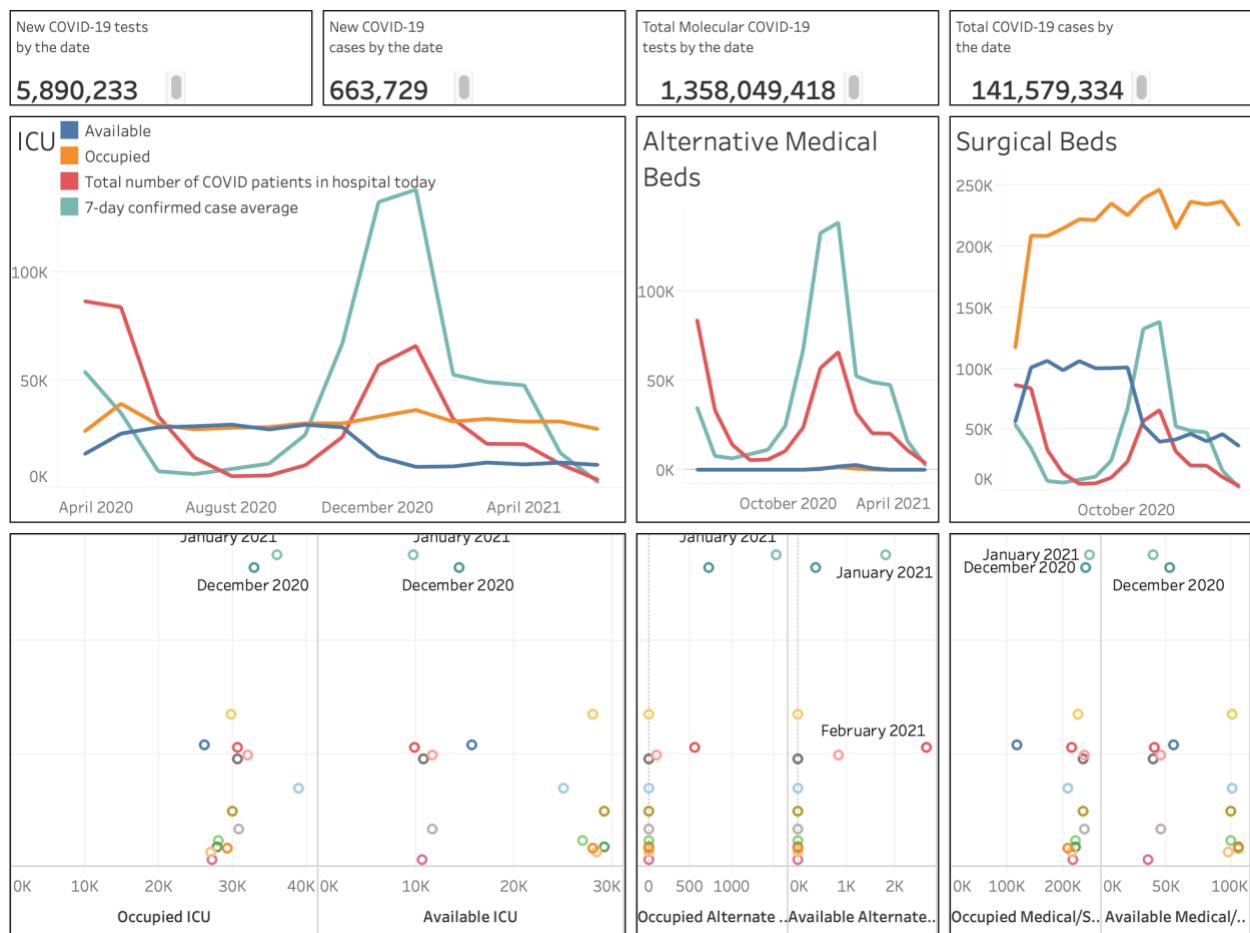Fig 14. Deep dive analysis of trend in Alternative Medical bed usage

Fig 15. Dashboarding of Hospitalization breakdown

KEY METRICS USED (DRIVER/OUTCOMES)

1. For our analysis, we have utilized a variety of characteristics. Primarily leveraged age, date, hospitalizations, and gender characteristics. We have tried to uncover the underlying pattern in COVID-19 data by answering a series of questions based on our findings and research objectives. The first dashboard elaborates the "Last 2 Week KPIs" which states the most affected age-group, most hospitalized age-group and most affected sex-group from which the UT Physicians can easily target the most affected group of people so that they can focus more or invest more to bring their curve of danger down.

2. We have used several very crucial variables for analysis in the second dashboard. Hospitalization and Death Rate are the main indicators for our visualizations in this dashboard. Races and Months columns had huge benefits for our analysis. We were able to analyze more deeper pattern between variables. Findings requires more analysis because data need to be free from bias to get most accurate results.

3. The key metrics used in the third dashboard that depict information regarding Cases vs. Hospitalization are the number of new positive cases, the total number of COVID patients in hospital today, the 7-day average of COVID hospitalizations, and the new COVID patients in the hospital that day. This helps the authorities to better prepare and predict for future hospitalizations based on the COVID positive cases and deep dive to analyze how patients are hospitalized.

4. The key metrics used in the fourth dashboard helps in analyzing the kind of hospitalizations that patients required during COVID and the trends in the usage of ICU, surgical and alternate beds. This information can help the hospital provide better resources in case of emergencies such as this recent pandemic.

13

CONCLUSIONS AND RECOMMENDATIONS

From the area curve of Positive new cases vs Estimated active cases, we concluded that Covid-19 cases started drastically decreasing from June 1, 2021. This could probably because of people getting vaccinated mostly in this period. From the Daily hospitalization curve, we can also see that after June 1, 2021, the hospitalization rates started getting low. We also concluded that people between the age of 20-29 are getting the most positive Covid-19 tests, people of age 80+ are getting hospitalized more often and the most affected sex group is Females. We recommend UT Physicians constantly check this dashboard for quick important insights so that they can target the group of people and prioritize their work.

The cases vs. Hospitalizations showed that the hospitalizations rate increased with the number of new COVID patients that had severe illnesses. Also, as the availability of ICU and surgical beds increased, the number of alternate beds were occupied significantly showing the patients being moved from severe cases to normal beds showing improvement in the treatment. This helps staff the hospitals in a better way in terms of resources and treatment planning.

Our aim to provide the system for Electronic Health Records to patients can use the above analysis and information to integrate the right set of details in every patient's record, assisting doctors as well as the infrastructure of the hospitals.