Analysis of Covid Cases and Mortality Rate

Mohammed Omar Khan (mok236)   **Ankit Sati** (as14128)   **Akshat Jain** (aj14128)

Table of Contents

Introduction

> The spread of Covid-19 has affected the whole world severely, despite many measures.

> We look into the various factors that impact the rise in COVID cases.

> We have provided various visualizations to show how each factor affects the rise in the cases.

> We also performed a time series analysis on the COVID cases reported from Jan 2020 to May 2022 and predicted the number of cases for 3 months after May 2022.

Objective

> To identify the nature of the pandemic and then to forecast its progression

> We want to analyse the different factors which affect the rise in COVID cases and then predict them using Time series Analysis.

Methodology

Datasets

> We have used multiple data sources to study different types of features which affect the number of covid cases.

> The idea behind using multiple datasets is to understand the different trends and effects of the rise in covid cases for example: hospitalization rate of adults, fatality rate in different states, etc.

Dataset

	date	state	fips	cases	deaths	fatality_rate%
55833	2022-12-04	Virginia	51	2153223	22582	1.048753
55834	2022-12-04	Washington	53	1859858	14739	0.792480
55835	2022-12-04	West Virginia	54	615332	7740	1.257858
55836	2022-12-04	Wisconsin	55	1929331	15684	0.812924
55837	2022-12-04	Wyoming	56	180925	1938	1.071162

Data Wrangling

> Data Wrangling is a key step because the data could contain unnecessary information, missing and inconsistent values

> In the case of visualizing the fatality rate for each state. We had cases and deaths for each state. We used both of these columns to calculate the fatality rate for each state as shown in the figure. This resulted in a much better understanding and visualization.
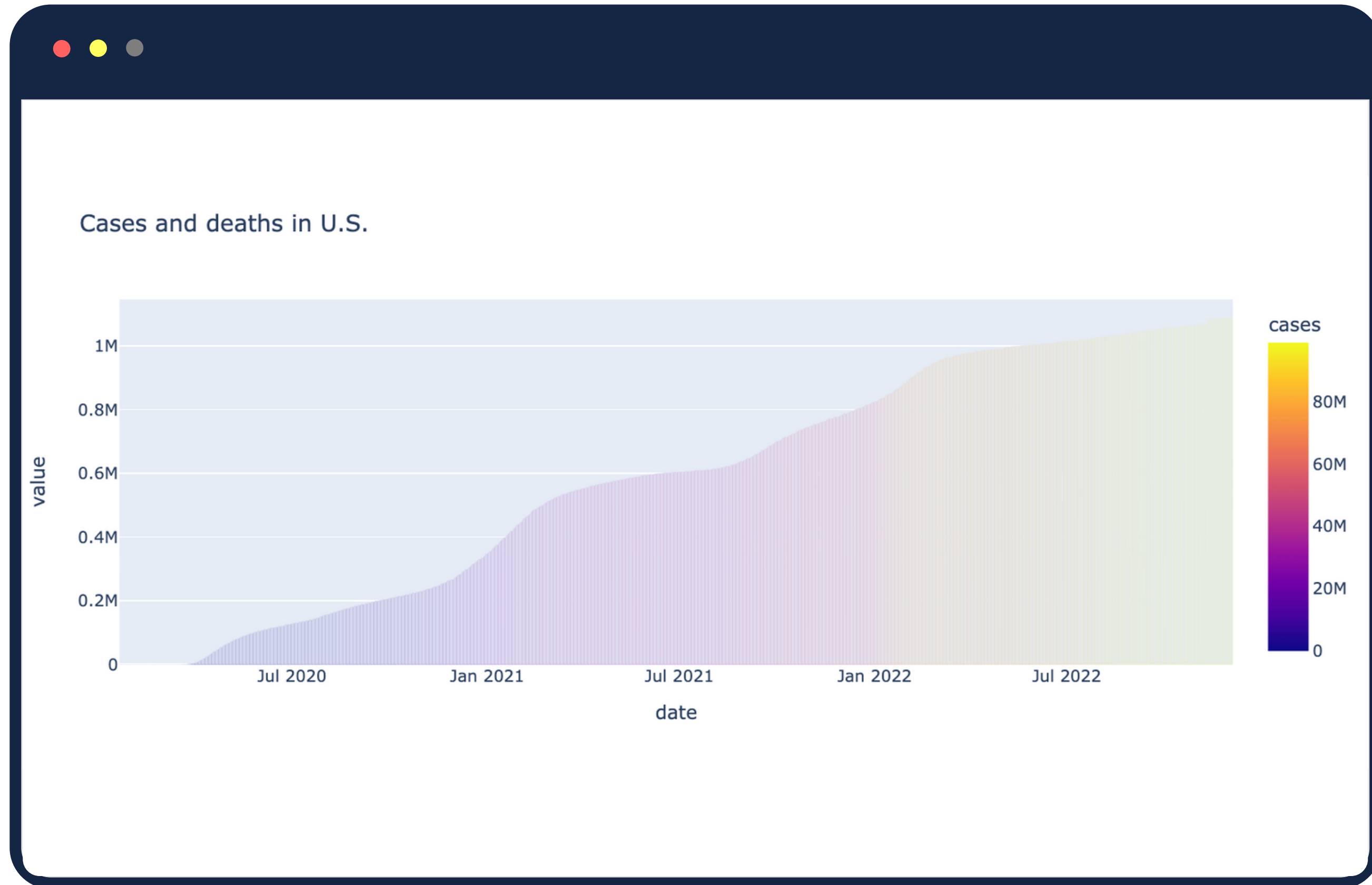
Data Visualization has a crucial role in understanding and analyzing the different trends in the data. Using line graphs, bar graphs and treemap, we try to analyze the different trends and factors responsible for the number of covid cases. Python libraries such as matplotlib, seaborn and plotly along with Tableau were used for creating the visualizations.
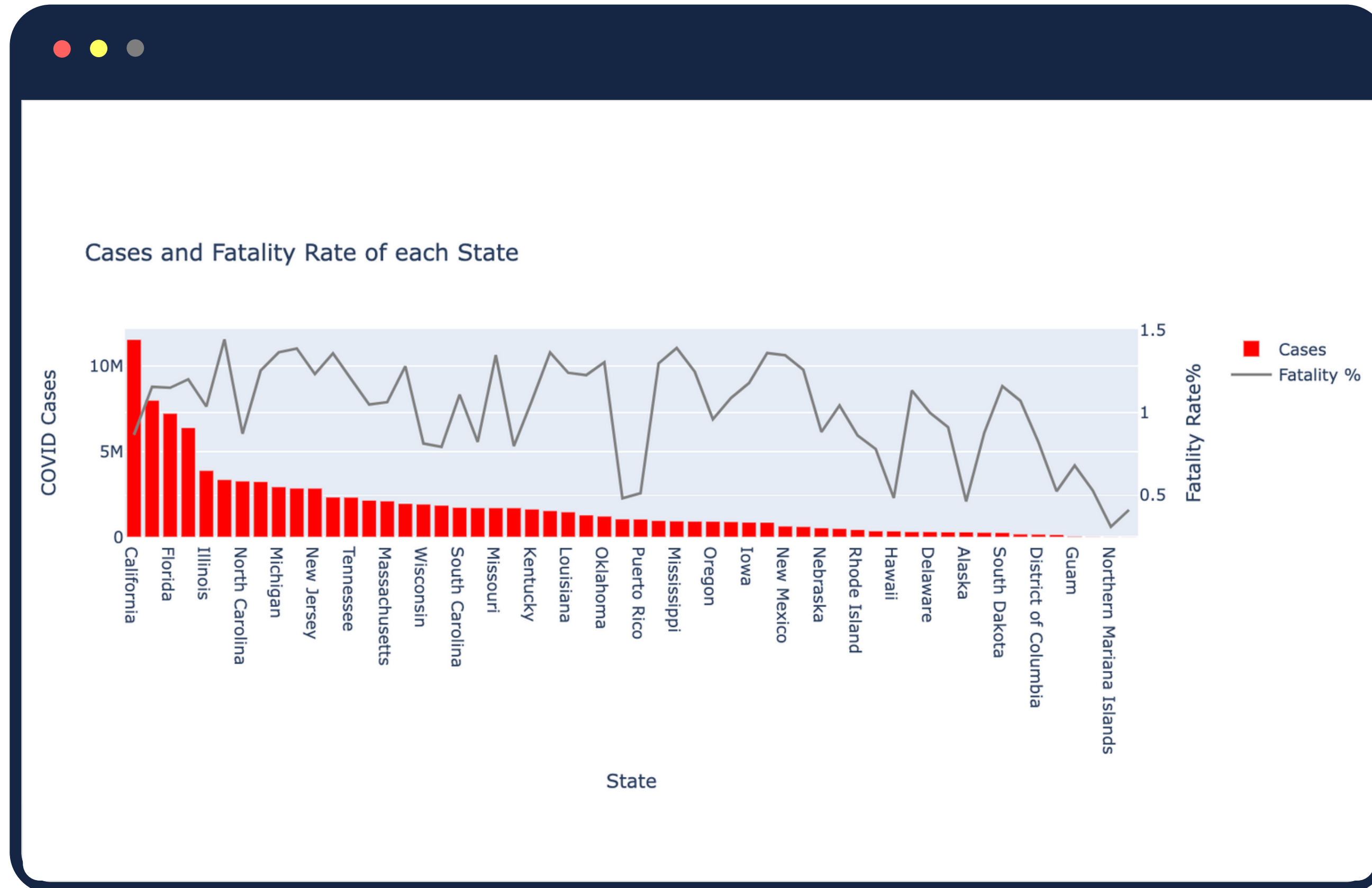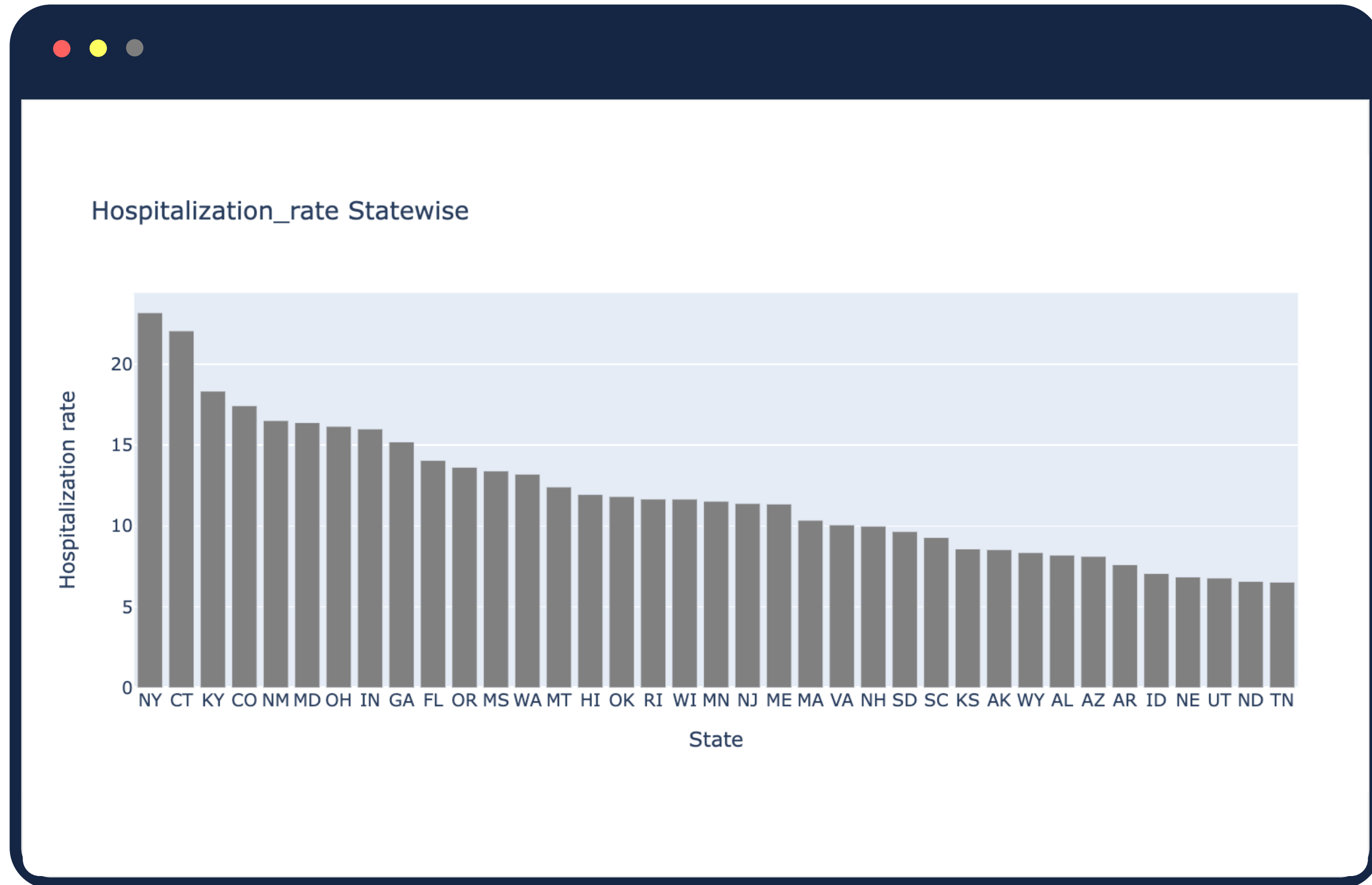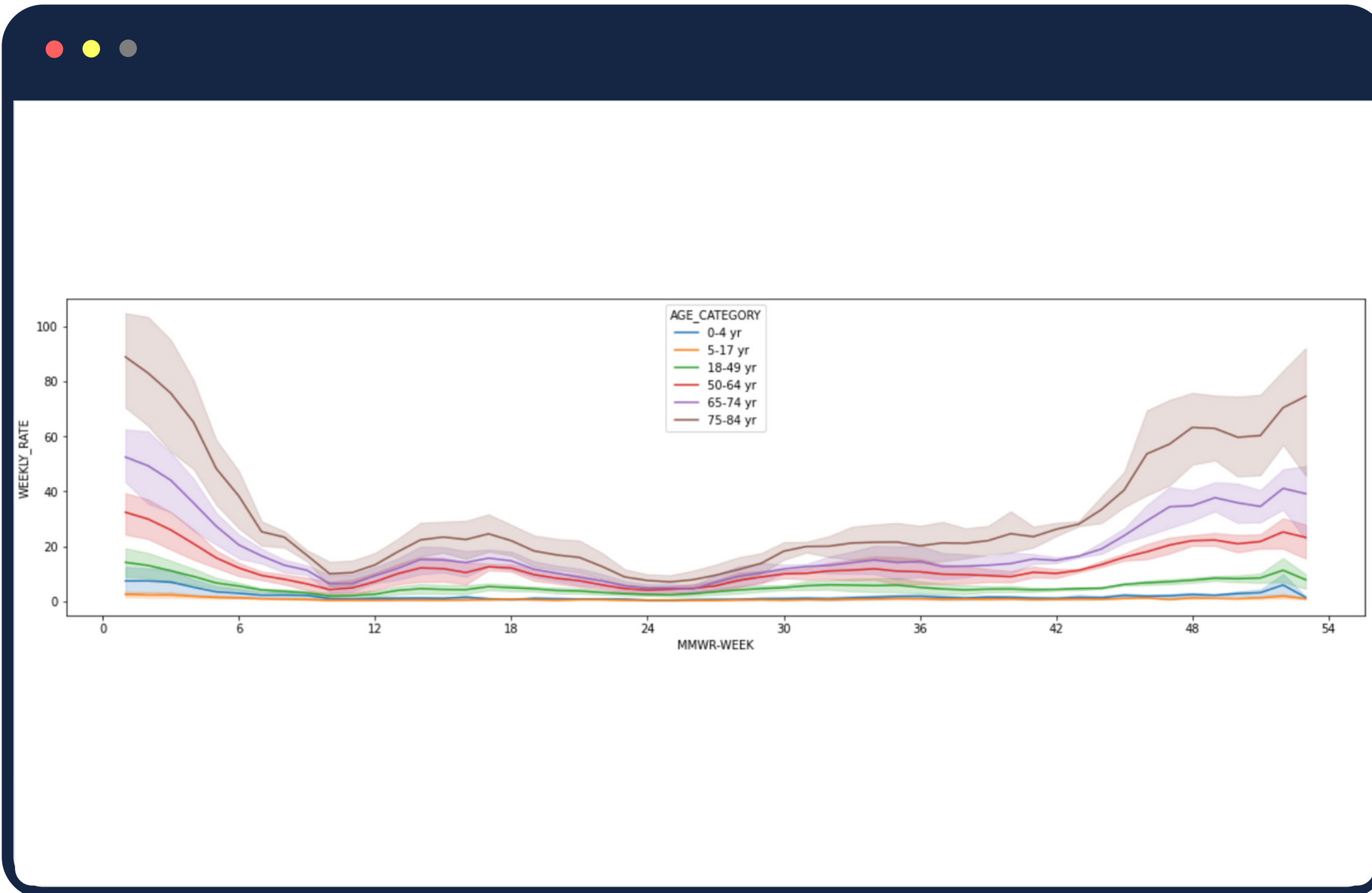
Visualizations

Visualizations

Visualizations

Cases and Fatality Rate of each State
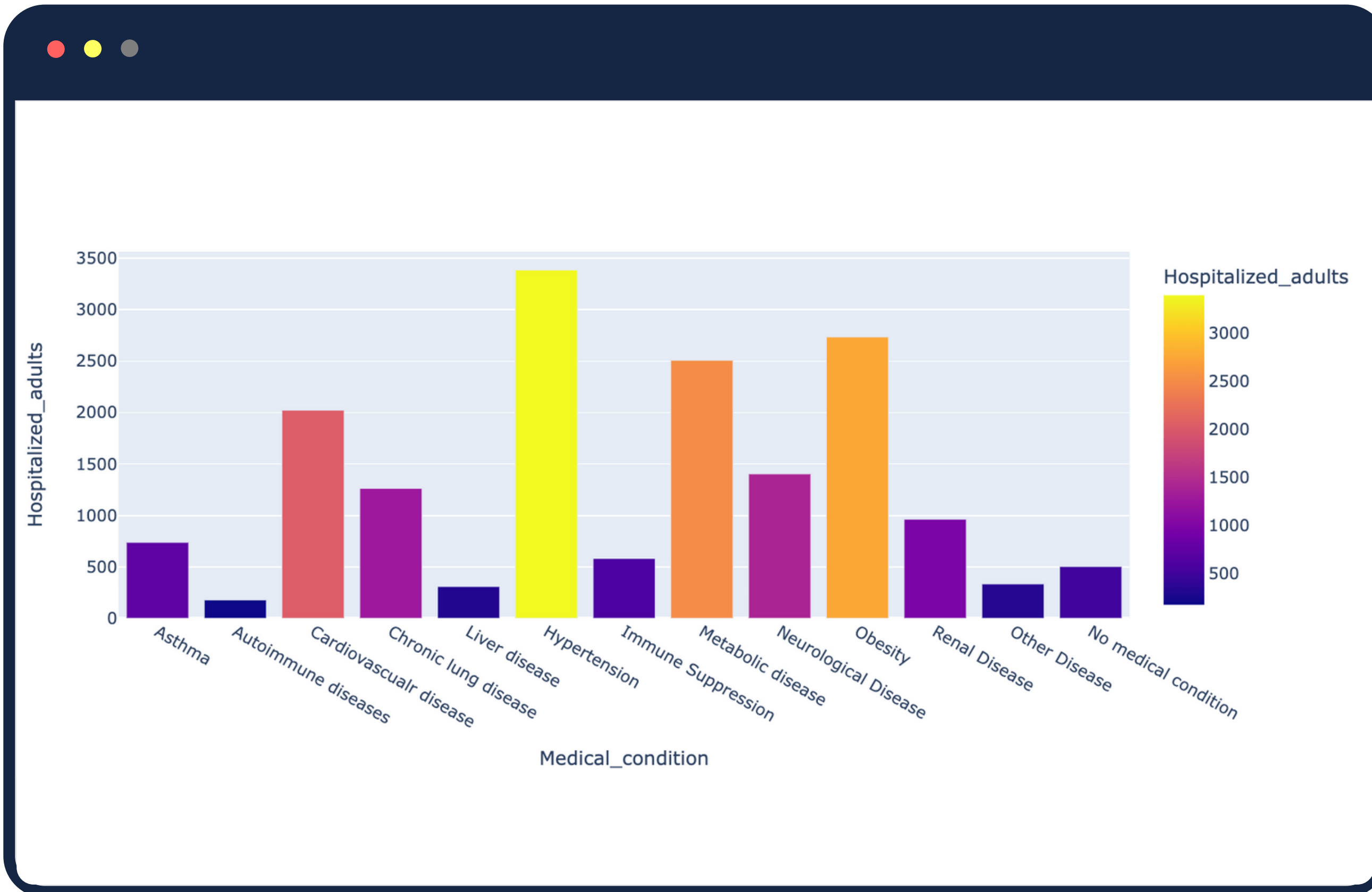
Visualizations

Hospitalization_rate Statewise

Visualizations

Visualizations

Visualizations

State wise total Cases and Deaths
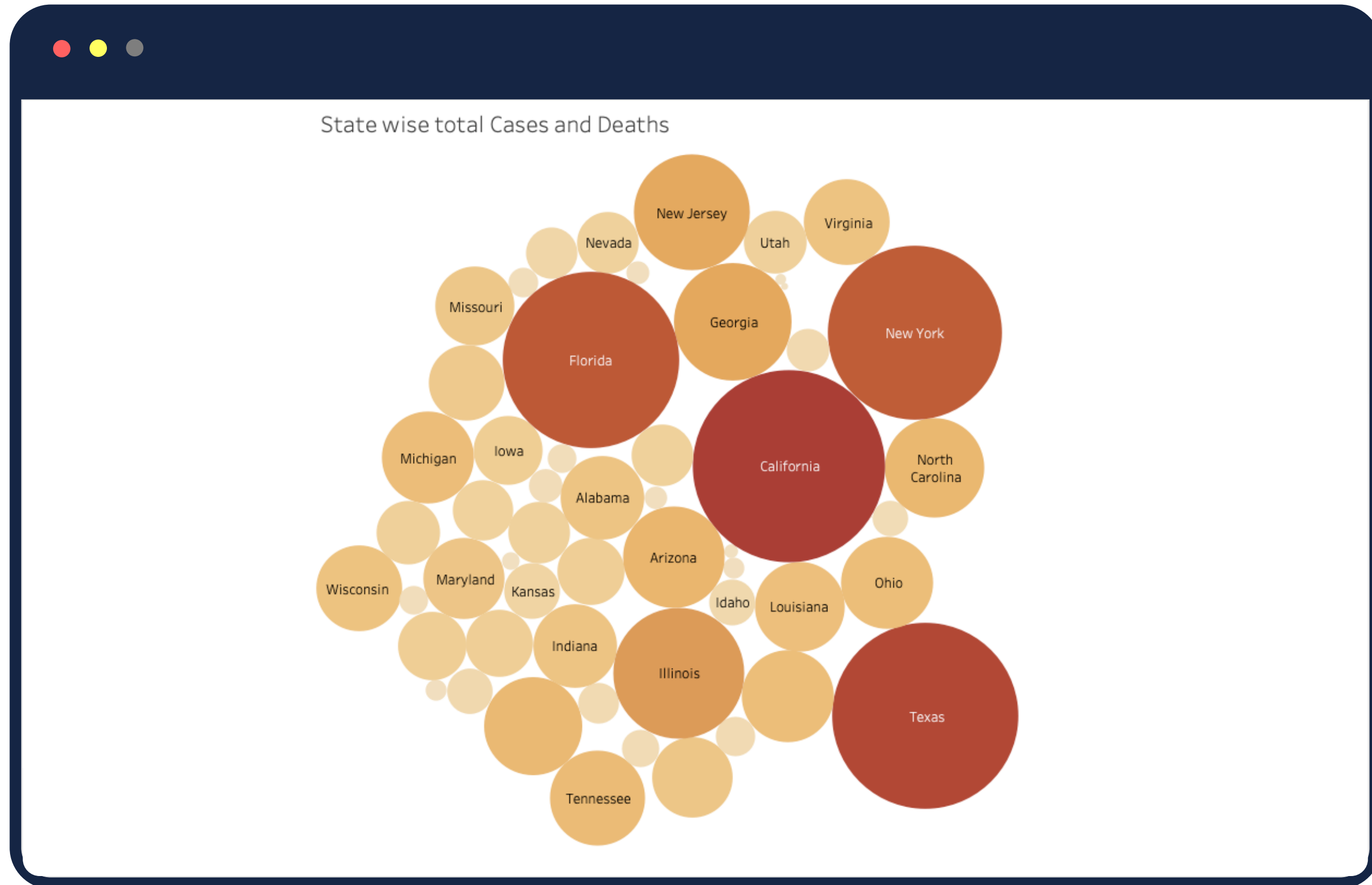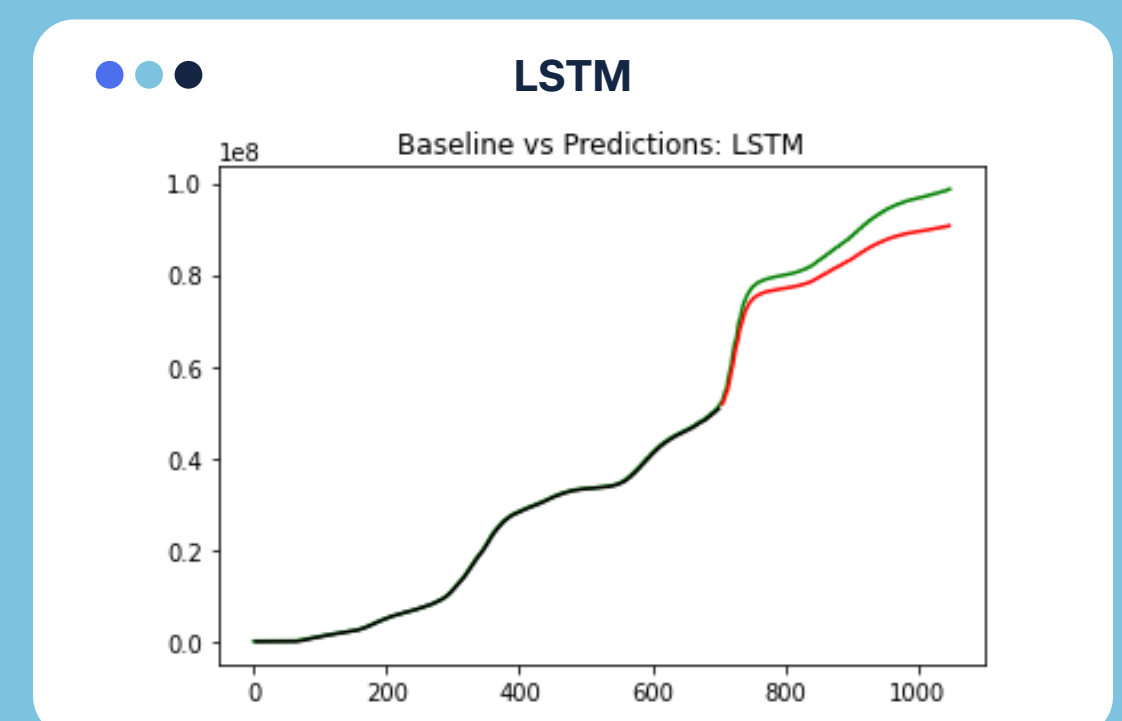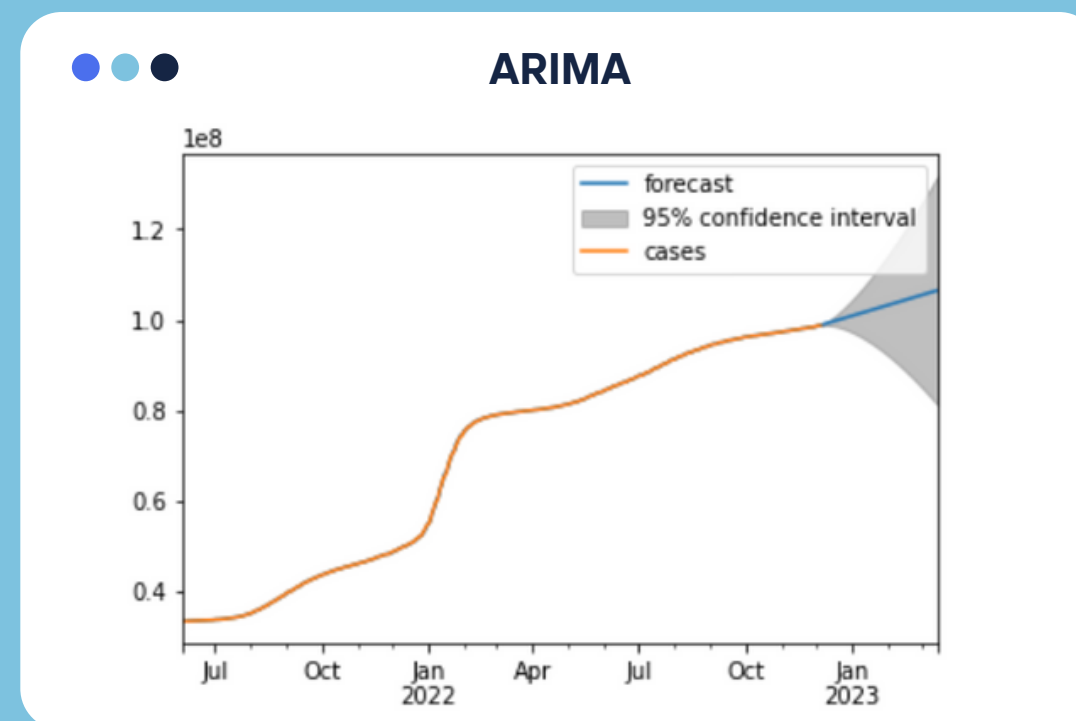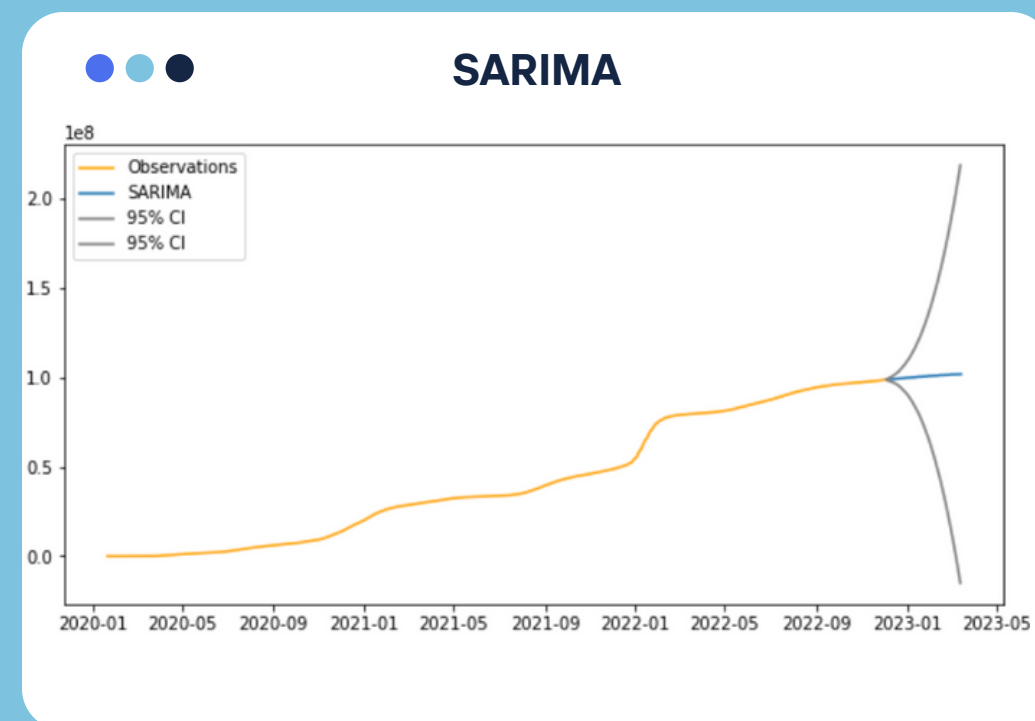
Results

> We worked with both ARIMA and SARIMA models and predicted the values for 100 days in the future.

> For both ARIMA and SARIMA models, we got an error rate of 0.00098. But the confidence interval keeps on increasing.

> This is opposite in the case of LSTM with a higher error rate but a smaller confidence interval .

Conclusions

These models are suitable for short time interval predictions of covid cases. or blur

Pandemics are far more complex than we think. There are many other factors such as virus mutations, vaccine production efficiency, population vaccination rates etc. that affect the accuracy of the model

Visualizations explained before indicated that different states as well as different segments of the population had varying effects of the virus

In the future, if such a pandemic occurs, then governments should be ready with strategies which would help these regions well before the situation becomes unmanageable

References

> Andrea Remuzzi, Prof, EngD, Giuseppe Remuzzi, Prof, MD COVID-19 and Italy: what next? - PMC, 2020 11-17 April

> Navid Mashinchi,Predicting number of Covid19 deaths using Time Series Analysis (ARIMA MODEL) | by Navid Mashinchi | Towards Data Science, Sep 22, 2020

> Sumedha Gupta, PhD, Archelle Georgiou, MD, Soumya Sen, PhD, US Trends in COVID-19–Associated Hospitalization and Mortality Rates Before and After Reopening Economies, June 25, 2021

> Bukhari, Syed Qasim, Massaro, Joseph M., D'Agostino, Ralph B., Khan, Sheraz , Effects of Weather on Coronavirus Pandemic, July 2020

> David W. S. Wong, Yun Li, Spreading of COVID-19: Density matters, Dec 23 2020

> Qiang Wang, Shuyu Li and Rongrong Li, "Forecasting Energy Demand in China and India: Using Single-linear Hybrid-linear and Non-linear Time Series Forecast Techniques", Energy Elsevier, vol. 161, no. C, pp. 821-831, 2018.

Thank you!