Carnegie Mellon University Heinz College
INFORMATION SYSTEMS • PUBLIC POLICY • MANAGEMENT

COVID-19 Vaccine Effectiveness and Forecasting of Mortality

Keerthi, Tejo, Sathiya, Siva, Hossein

Objective

Develop an AI model to predict COVID-19 death rates based on historical data.

Approach

Phase 1: Analyze data from May 2021 to June 2022 using ARIMA, SARIMA, and LSTM forecasting models.

Phase 2: Extend analysis to September 2022 and consider Multivariable Time Series Models, including vaccination patterns' impact on death trends.

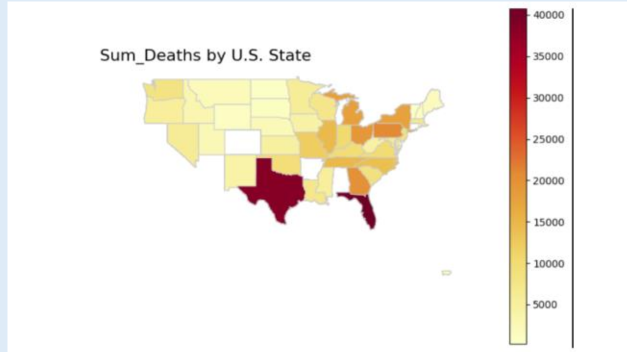
Data Sources

1. Johns Hopkins University Data: FIPS code, demographic info, daily death counts.
2. CDC Data: Vaccination dose info, state, and county information.

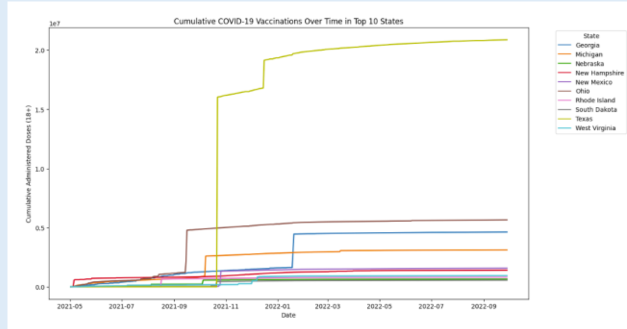
Methodologies

- Implemented models like SARIMA, Random Forest, and LSTM for time series forecasting.
- Analyzed patterns to deduce affects.
- Utilized MLflow to track experiments, log model parameters & Dockerized the env.

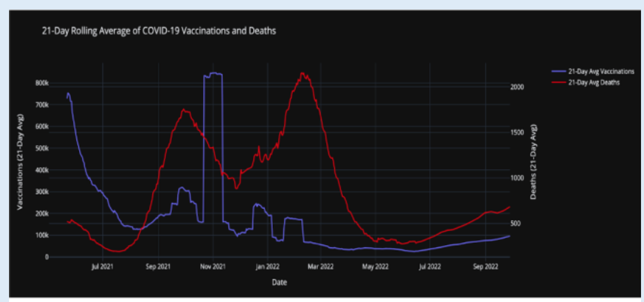
Exploratory Data Analysis



Choropleth indicates the total number of COVID deaths per state. Of the data available in the plot, the most affected areas were Texas, Florida, and Pennsylvania.



This is a step plot displaying the cumulative number of COVID-19 vaccine doses administered to individuals aged 18 and over in the top 10 U.S. states. There are steep increases in the cumulative doses administered at certain points, particularly early in the time series.



The graph shows a 21-day rolling average of COVID-19 vaccinations and deaths over time. The blue line, which represents the 21-day rolling average of vaccinations, shows an initial peak followed by a decline and then several smaller peaks. The initial peak may correspond to the initial vaccine rollout when there was high demand and eligibility was expanding.

Insights

When the model was trained with data up to Sept 30, 2022, it showed predictions that closely aligned with the actual test data. This was attributed to a pattern shift beginning around March 2022, which was captured in the training data. Training the model with data only up to February 2022 led to less accurate predictions. The model missed capturing new patterns that emerged after February 2022, showing that it couldn't adjust to the evolving trends not present in its training data.

The insights gained from this project are likely valuable for public health analysis, policy-making, and understanding the impact of COVID-19. The emphasis on the importance of recent data in model training is a critical takeaway, highlighting the need for adaptability in predictive modeling during rapidly changing situations like a global pandemic.