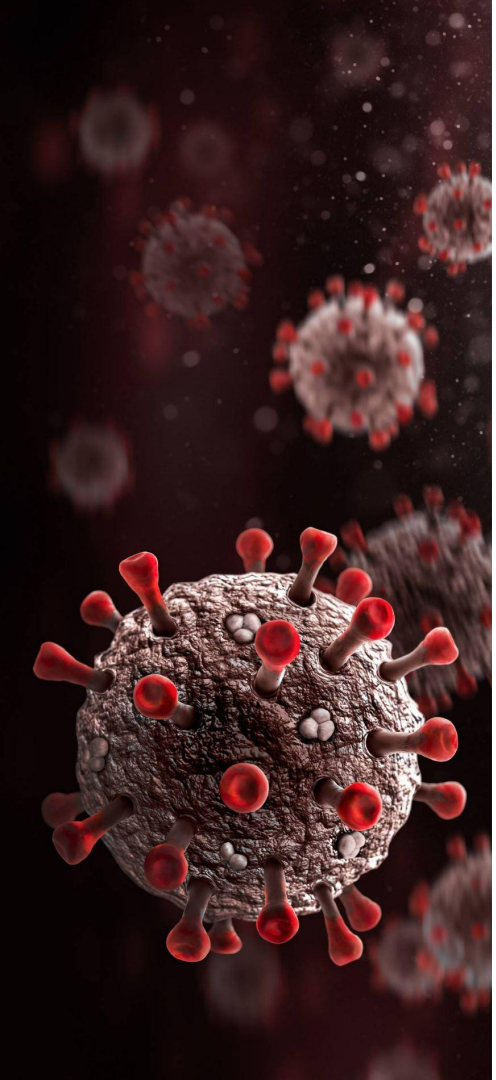
Drawing Insight from COVID-19 Data,

Informing Resource Allocation Decisions

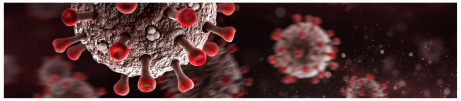


Problem Statement

COVID-19 has been with us for 2 years now, with most people fully vaccinated and boosted in most states, but less so in others.

- Insights from data on covid-19 vaccination, surveillance, and death rates
- Identify states more susceptible to new cases
- Attempt time series model
- Inform resource allocation decisions

https://hub.jhu.edu/2021/01/15/sars-cov-2-covid-19-variants-new-strains/

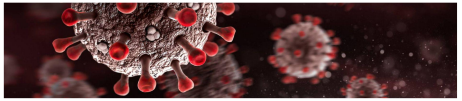


Key Questions

1. To which states should we distribute additional covid anti-viral therapeutics and focus outreach campaigns?

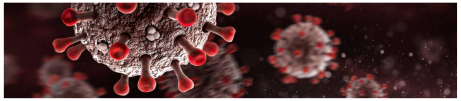
2. Can time series and classification models offer more granular insight?

https://www.centennialco.gov/Residents/Health-Environment/Coronavirus



The Data

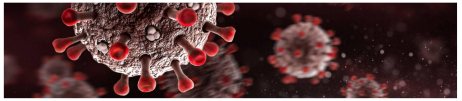
- COVID-19 Cases and Deaths by State
  - Confirmed and probable cases and deaths
- Vaccination Trends by State
  - Breakdown by percentage of people vaccinated
- Surveillance Information by state
  - Demographic data related to hospitalization rates
- Travel Data
  - Trips greater than 500 miles



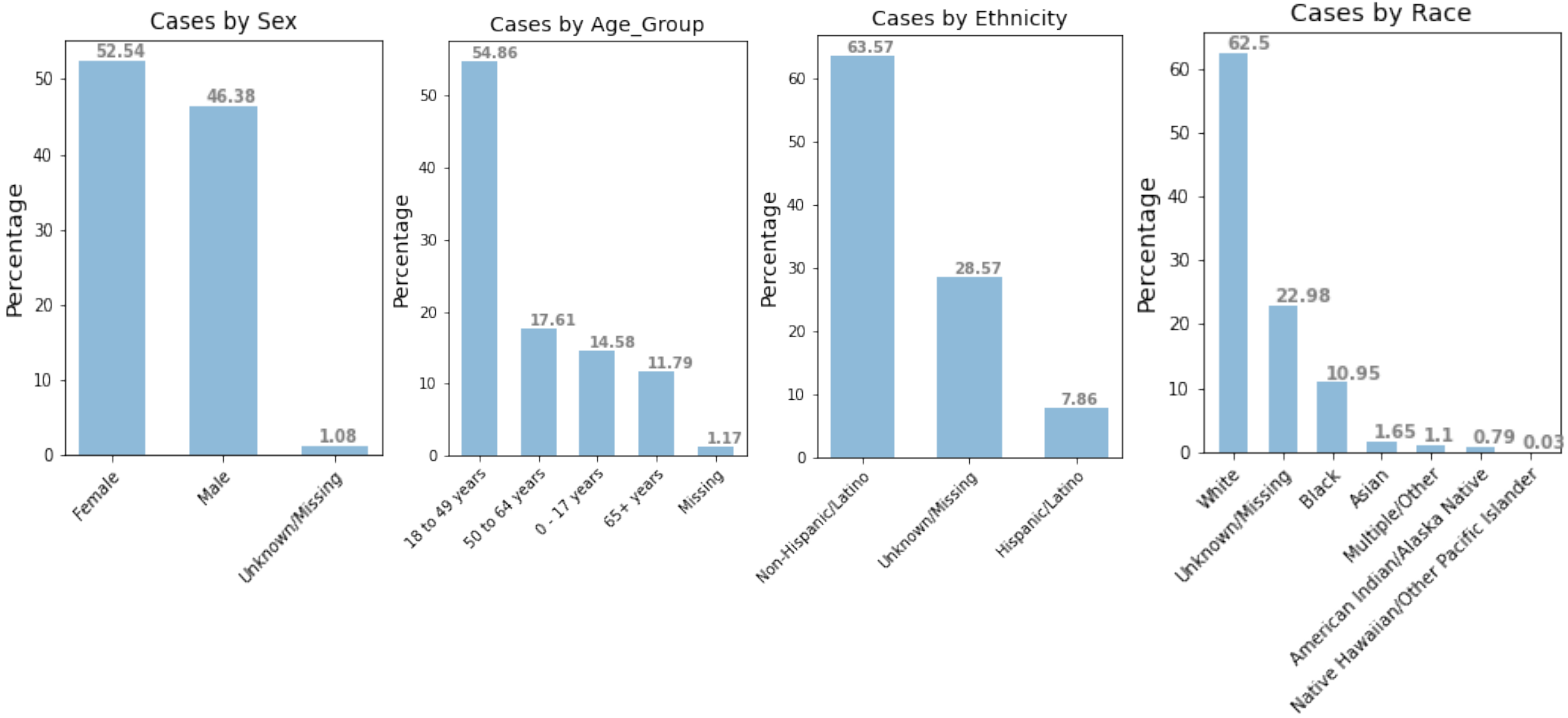
Approach

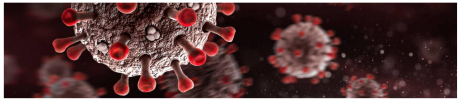
Criteria to identify states that might need additional resources:

- Monthly new cases in each state
- The percentage of people in a state with both a primary series and booster

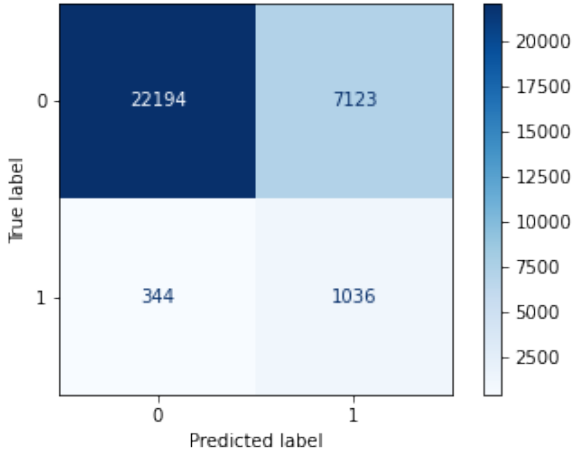


Demographic EDA- COVID





Classification model- Hosp prediction and inference



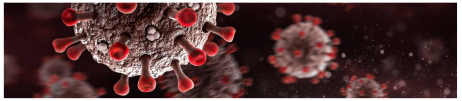
factor	exp_coef
age_group_65+ years	27.285825
age_group_50 to 64 years	6.799703
res_state_NJ	3.955109
res_state_KS	3.639672
res_state_KY	2.440233
month_3	2.328145
month_5	2.322213
res_state_NV	2.310894
age_group_18 to 49 years	2.257014
res_state_OK	2.098832

-Logistic Regression with Under Sampling

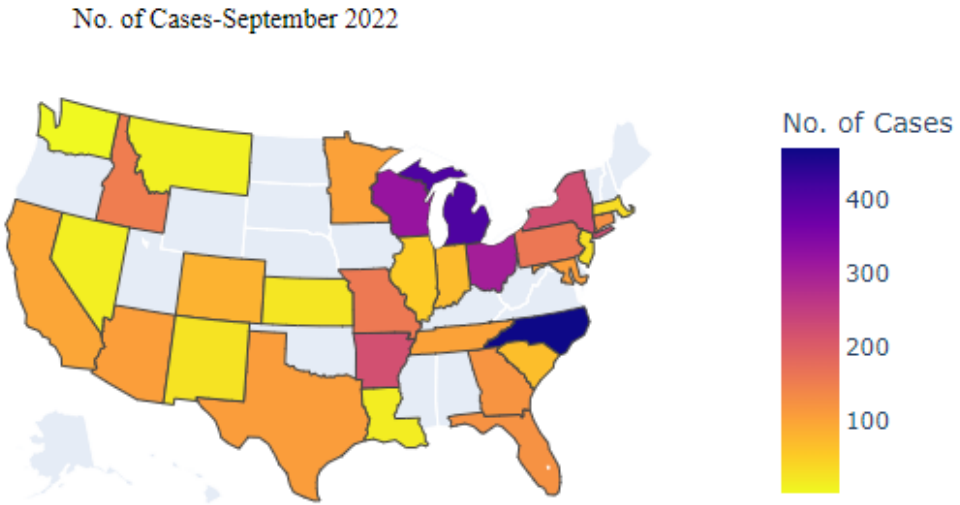
-Balanced Accuracy: 0.75

-Recall: 0.75

-Misclassification rate=0.24



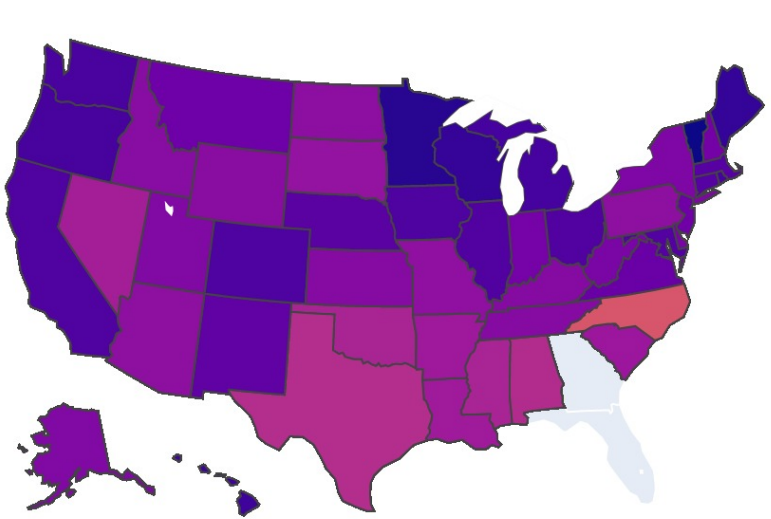
No. of Cases - September 2022

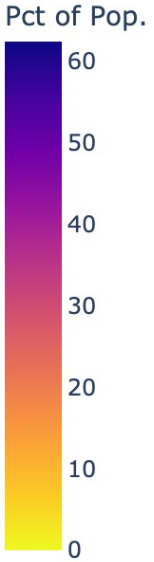


States with highest no. of cases

State	No. of Cases (Mean)
NC	471.0
MI	409.0
WI	321.0
OH	303.0
NY	225.0

Pct of Pop. with Completed Primary Series and a Booster, June 2022





States with Lowest Pct. w/ Primary Series and Booster:

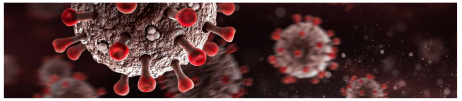
North Carolina - 28%

Alabama, 37%

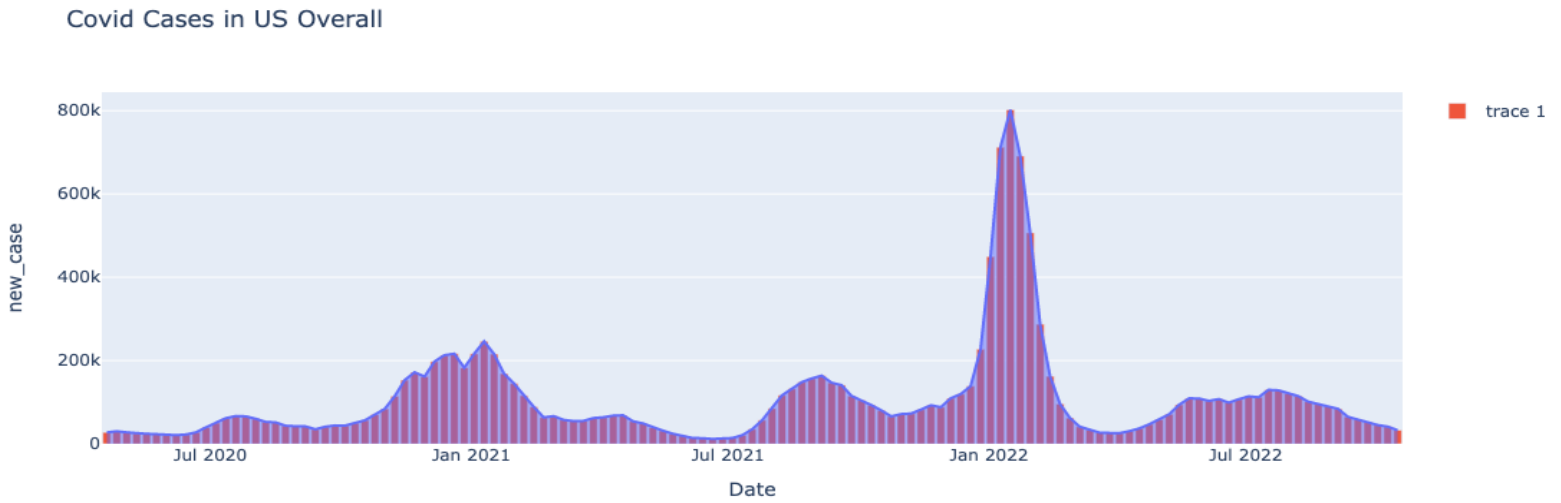
Oklahoma, 38.6%

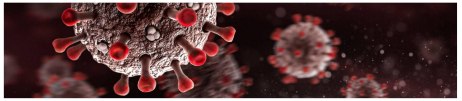
Mississippi, 38.8%

Texas, 39.2%

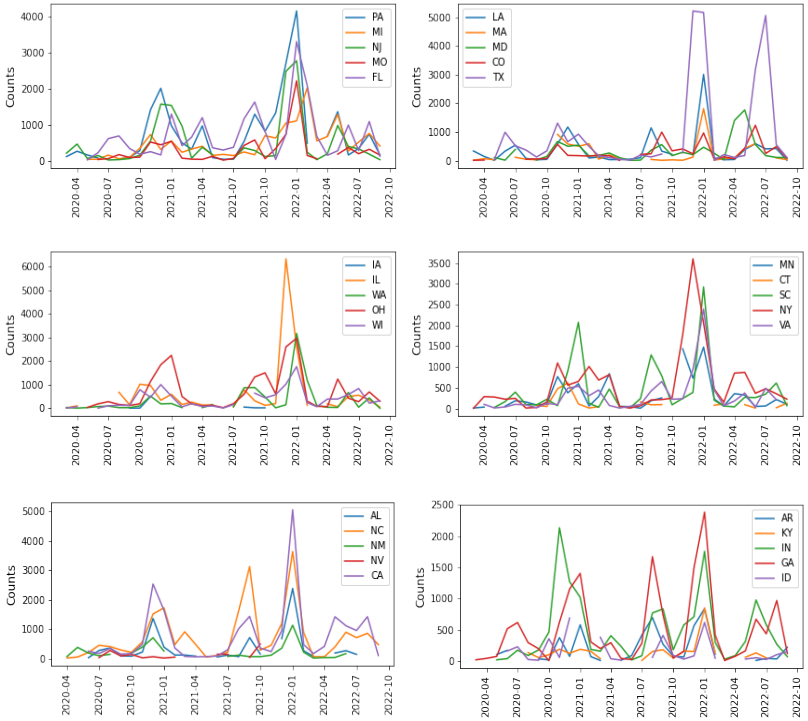


Covid Cases Evolution in US

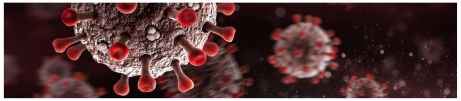




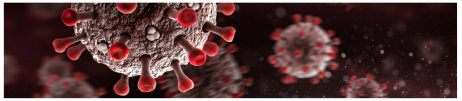
Covid Cases Evolution by State



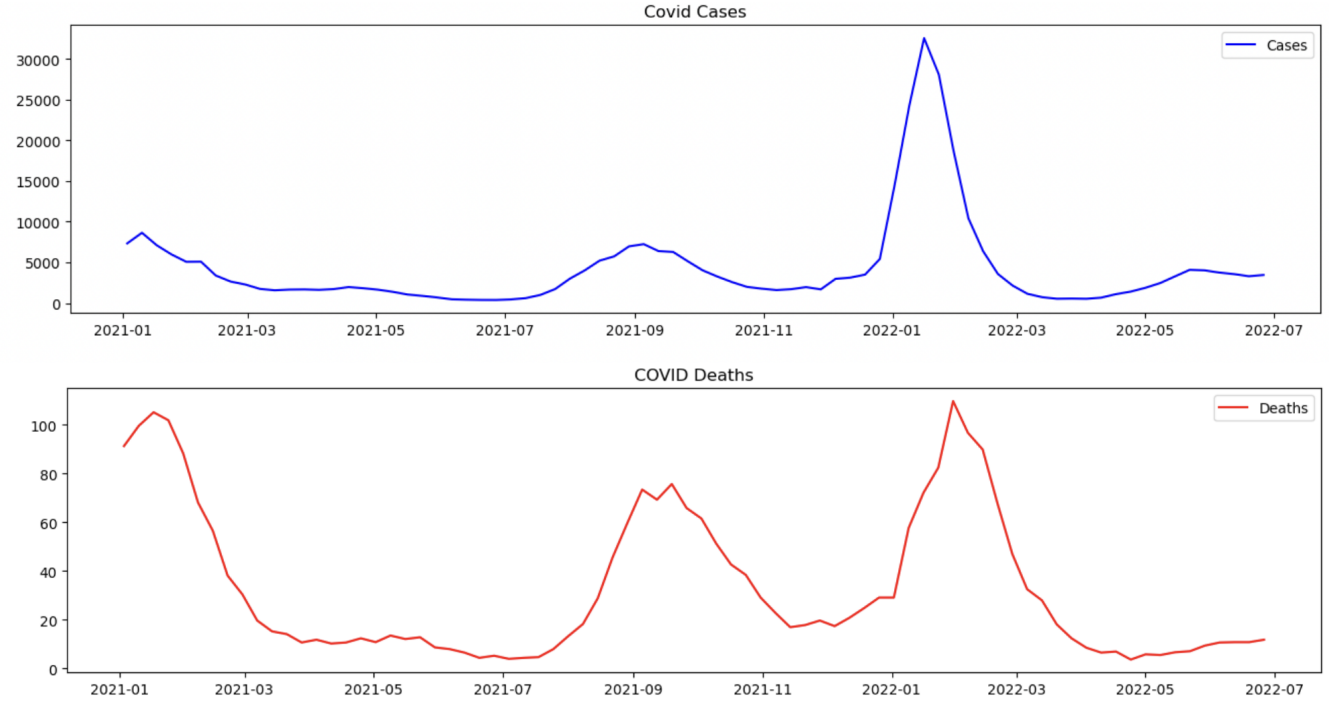
- Major peaks during winter 2020 and 2021 each state
- Consistent peak across states in summer 2022 (April-August)
- Some of the states show higher number of cases until September

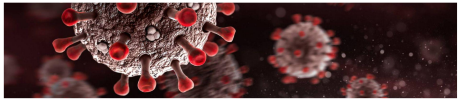


Covid Cases Data



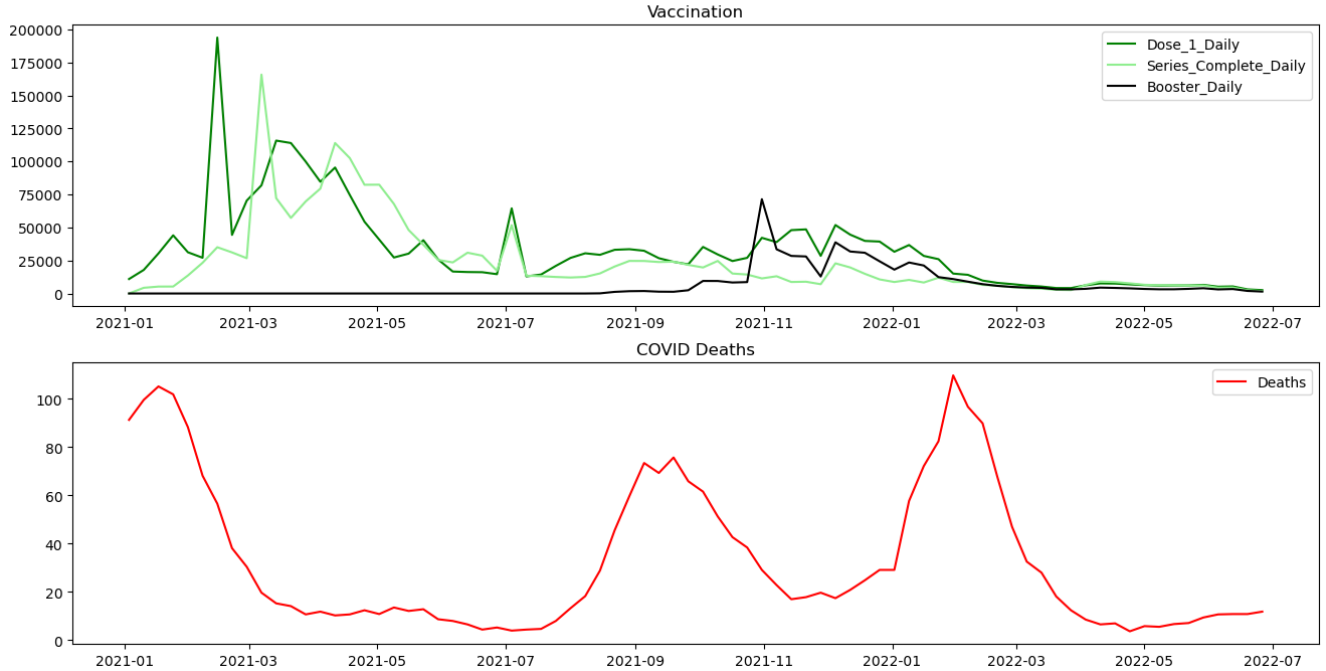
Covid Cases and Deaths Data Insights

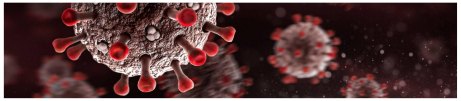




Vaccination and Deaths Data Insights

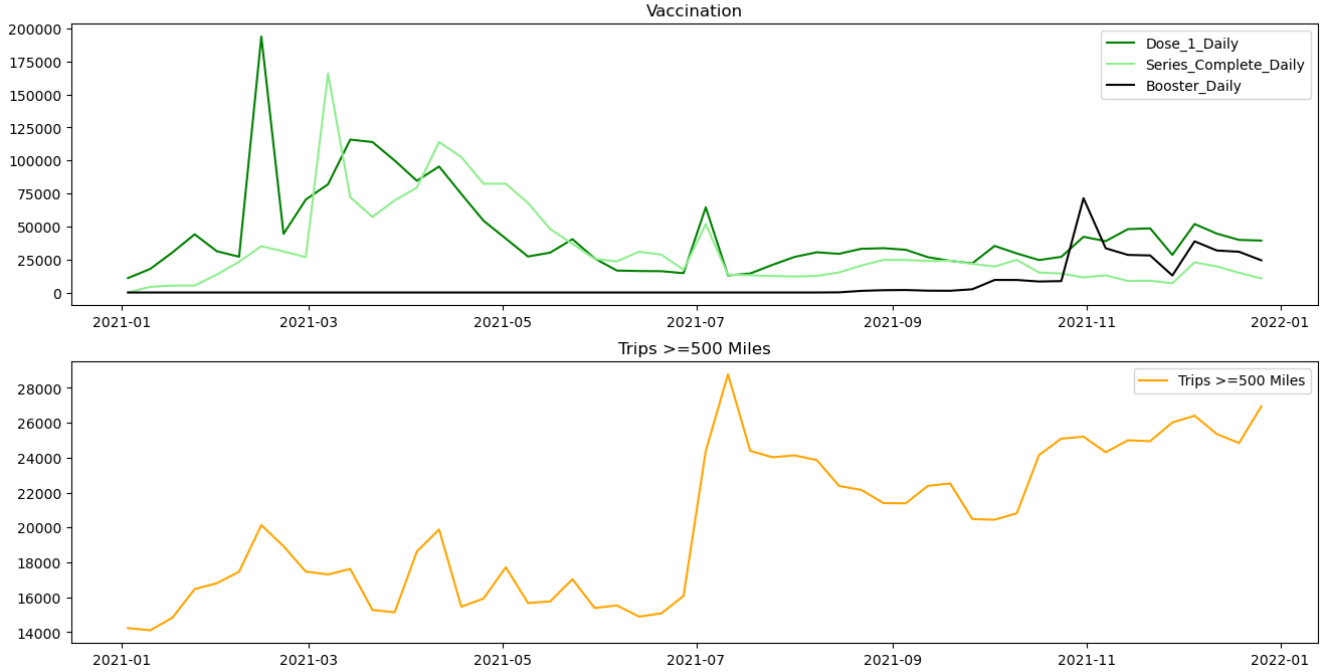
North Carolina Vaccination and Covid Deaths Over Time



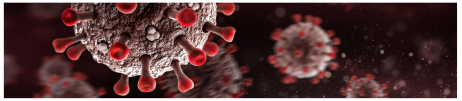


Vaccination and Travel Data

North Carolina Vaccinations and Population Not Staying at Home for Trips >=500 Miles

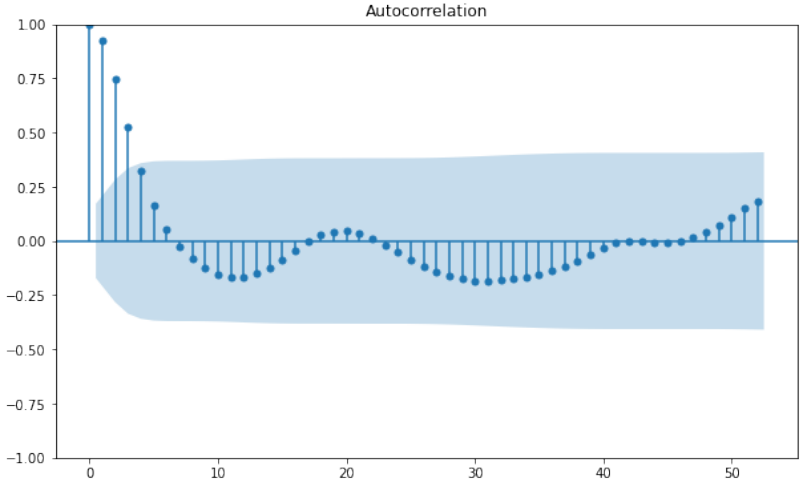


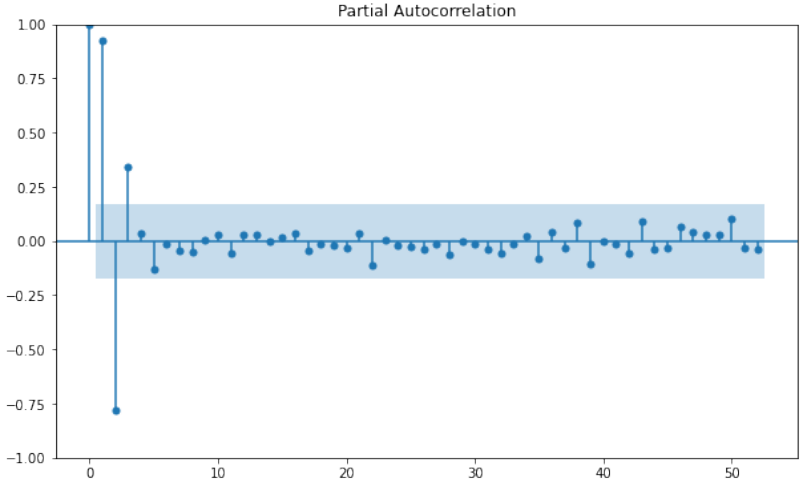
e

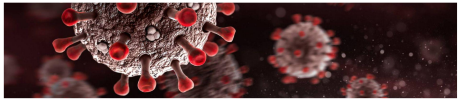


Time Series Model for Cases in US

- High autocorrelation on the 3 first legs
- No trends/seasonality was found at the data

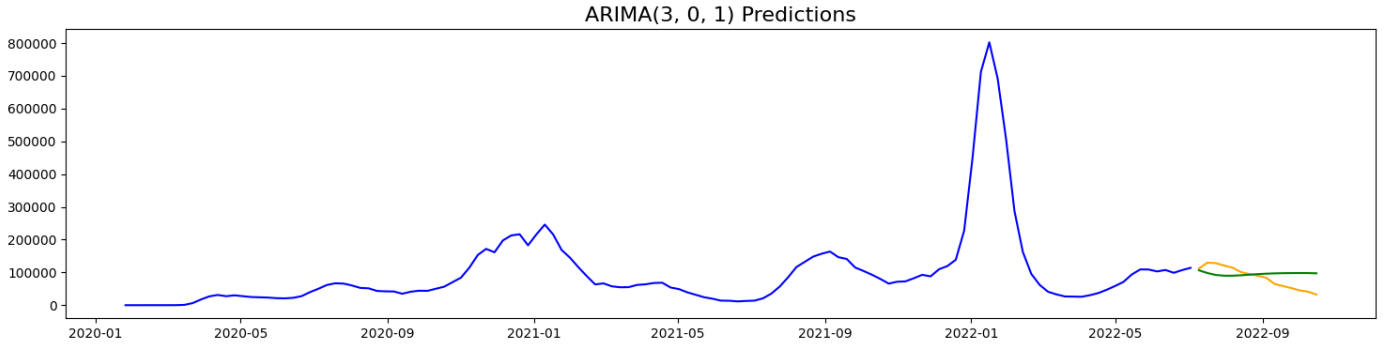


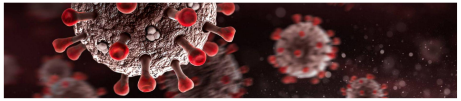




Time Series Model for Cases in US

ARIMA	(0,1,0)	4237225928.63
ARIMA	(3,0,1)	2791954341.41
SARIMA	(3,0,1) x (1,1,1,22)	3638844274.47
VAR (North Carolina)	covid_deaths, covid_cases, trips	20630275.43



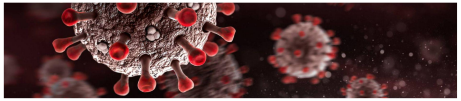


Findings/Recommendations

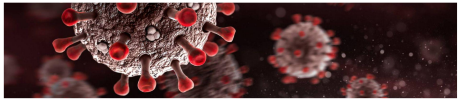
- Focus on North Carolina:
  - Highest number of new covid cases over the past month
  - Lowest percent of people w/ primary series and booster (28%)

- Time series modeling is difficult:
  - Tried multiple models: ARIMA, VAR, SARIMAX
  - Lack of domain knowledge

- Imbalanced classes affected classification model performance:
  - Able to identify factors that contributed to hospitalization
  - Collect more data to balance



Thank you!!



Demographic EDA- COVID