COVID-19 Case Study

Introduction and Project Overview:

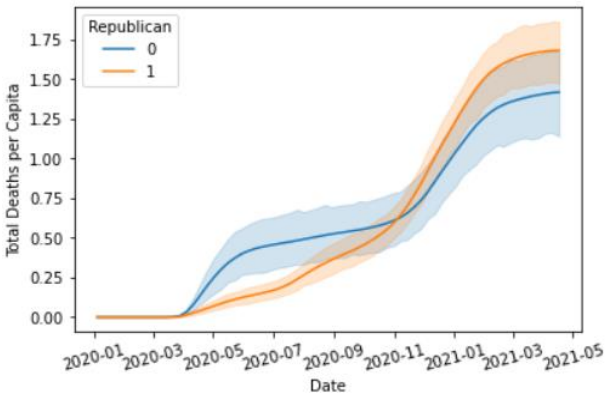
The Covid-19 pandemic has changed the landscape of the world since its global outbreak in the late months of 2019, through both the devastation of the death count but also in the economic effects that shall be felt for years to come. The purpose of this project was to identify factors in the United States that resulted in a greater number of deaths per capita. These factors were per state. Below is a brief review of this project.

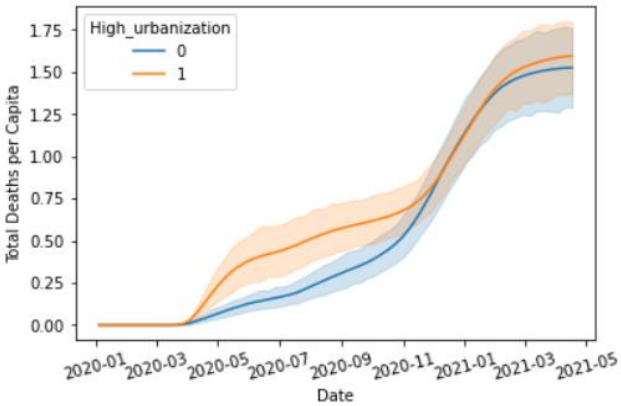
Data:

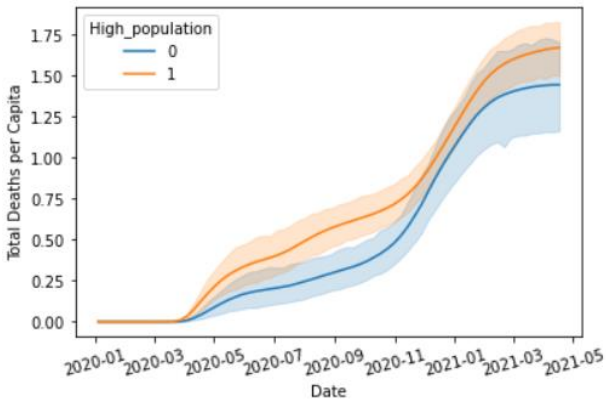
The data used is available on the CDC and the US Census Bureau's website and is free to access. The 5 primary state attributes that were looked at were as follows: GDP, voting patterns, population size, obesity rates, and the degree of urbanization. Voting patterns in this case were defined as how the state had voted for the past 3 presidential elections, democrat or republican. Obesity is defined as a BMI of over 30 and was a percentage number per state. Urbanization was defined as any incorporated place with a population of greater than 2,500 people. This number was the percentage of state population that lived in an urban place.

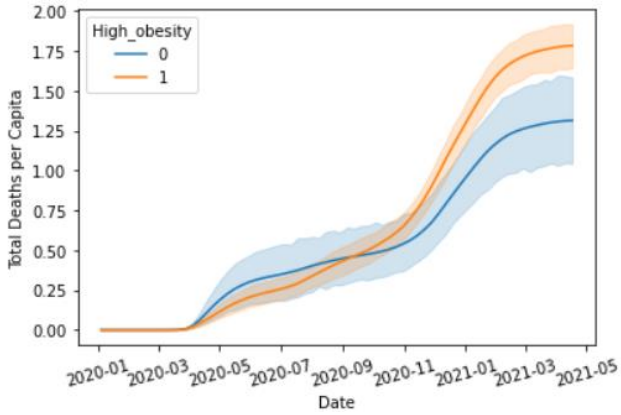
Graphs:

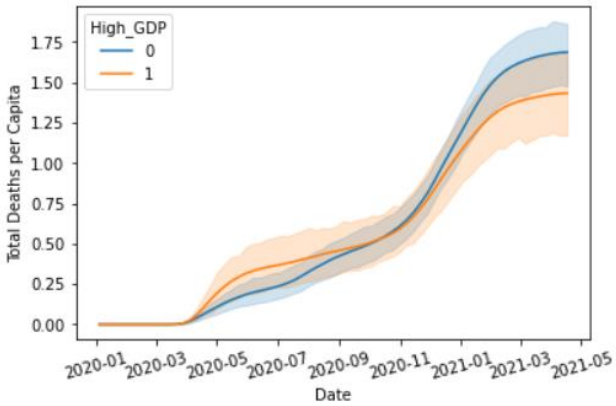
In the graphs below, the shaded region for each color is 2 standard deviations with the line representing the mean of the data.











Results:

To sum our findings, it looks like at the time of this project:

-The states with higher rates of obesity have a far higher death toll.
-Republican states have a higher death toll.
-The states with higher population have resulted in a higher death toll.
-The states with higher urbanization have not resulted in a higher or lower death toll.
-The states with higher GDP have resulted in a lower death toll.

While this project was more about data visualization than modeling, a model was created to see if the current features were enough to predict the total number of deaths. A linear regressor (LR) was used vs a random forest regressor (RF). The RF performed significantly better but still produced a negative R squared. Overall, the model was a failure. This was not unexpected though, as there was not an incredible amount of data. Considering the goal of this project was exploratory and not predictive, it can be said to be a success in spite of its modeling.

Future Directions:

There are, hypothetically, an infinite number of features to explore. One could see how differing policies or climates or regions affect COVID deaths.