TO achieve our analytical goal, we merged 7 datasets into a master dataset by state and by county for data investigation. There are over a dozen features from these dataset but we picked 8 features for our final model.

Here in the following presentation, we present most of our finding at the state level. We QC’d our models at county level.

The target feature is the vaccination rate, which is published by CDC and we cut off the data at July 31st, 2021. The vaccination rate is defined by people that are fully vaccinated divided by the total population of the state.

Then we used a dataset the CDC published earlier this year that surveyed people from the entire state on how hesitate they are about getting covid vaccine. The answers range from no hesitation to high hesitation. We summed up all different degree of hesitation to a new index as total hesitation.

And in February this year, CDC also did a prediction of vaccination outcome based on some county wide medical related concerns. In this study, we want to check these prediction against real world outcome to see if using these themes can explain the actually observation. These themes include Historic under vaccination such as the percentage of kids of certain age got their MMR, TDap vaccination. CDC is using that data to predict covid outcome. They also used social demographic barrier as a predictor, which include the minority population in the community and the income level, etc. They also assessed the healthcare seeking behavior of the community, that includes the location of the hospitals and how well the public transportation systems were set up and available. They also investigated people’s medical behavior such as regular doctor visit and annual check ups.

So they come up with the index of concern on each theme, the higher the number the higher the concern. There is a index called CVAC, which is a weighted average of the 5 themes.

In addition, we used to the 2020 presidential Election outcome to benchmark the political view of each state, and used the religious index and percentage of people believe in god to indicate how religious people are in each states.

In this exercise we want predict the vaccination rate using all of the predictors. And by finding the ‘good predictive model’, we want to answer the question: What are the most important factors that limit our vaccination rate?