Matthew Shinder

Capstone Two - Project Proposal

COVID-19 is a contagious disease that was caused by the SARS-CoV-2 coronavirus. While first known to originate in Wuhan in December of 2019, the virus quickly spread around the world creating a global pandemic.

The goal of this project is to look at trends of COVID-19 data, specifically inside each US county. We can first find trends between counties and see which counties had an uptick in covid cases whether it be in short surges or over a long period of time. We can then look into these specific counties and tie in demographic data such as age, wealth (economy), political affiliation, location, etc in order to find potential correlations.

We will use the US counties COVID 19 dataset (from Kaggle) which contains COVID-19 case numbers dating from as early as February of 2020 and up to May of 2022. After data wrangling and modeling our COVID data, we can look into the US Census Bureau dataset to get demographic information by county.

To solve this problem, we will build models upon the COVID 19 dataset and then attempt to draw correlations with the demographic datasets based on each US county and display any meaningful results through graphs and tables. Specifically, we are looking at US counties that either had short surges (COVID cases in a short period of time) as well as counties that had a lot of cases over a long period of time (based on population density).