Final Project DSC530

Ryan R. Gough

My original hypothesis was to ask the question whether covid testing and positive covid rates suffered from the class size paradox. However, as I began working the data to manipulate it. I realized that this was an unattainable goal. The dataset in which I was working did have a variable for positive covid rates. Which I assessed as the total\_cases, as these would only be verified by an actual test. But, when I parsed through the data, there was no metric for the amount of people actual tested. This was not collected within the dataset. I had to change my hypothesis to find actionable data. Therefore, I sought to see the difference between confirmed covid cases and covid deaths. That there would be a difference between nations, as governmental interventions varied between countries.

I found that for the country of Russia a confirmed diagnosis of covid had a predicted death of near 99%, while in the United States it was lowered to 40%. However, there is much more data that I needed in order to clean the differences between these. As we know just because someone has covid does not mean they all die. I believe this my EDA would of performed better with a better data profile. I do believe that if I had a variable that measured how many people were tested, as well as the amount of people who had covid. I would be in a better place to get the results I originally intended. The assumptions that I made with the data, I think is fine. However, given what I had, I didn’t have much else of a choice. Some of the challenges I faced were to me incomplete data profiles from countries that are known to not have reliable data. Thus, analyzing those countries resulted in questionable results.