**DS1 Final Report**

**Anandi Gupta**

**Word Count: 3027**

**Introduction**

The Covid-19 pandemic has devastated the incarcerated population in the U.S., infecting over 400,000 prisoners and killing over 2,600 prisoners to date.[1] It is estimated that one in three inmates in state prisons were infected by the virus as of March 2021.[2] Overcrowding, low testing rates, poor sanitation, and limited access to personal protective equipment have made prisoners particularly vulnerable to the disease.

For my project, I compare infection rates and death rates among incarcerated populations in state and federal prisons and the general population by county to highlight the disproportionate impact the pandemic has had on prisoners. I further examine the extent to which county level demographic and economic characteristics can predict the gap in infection rates between the incarcerated population and the general population. In the first section of this report, I provide a problem statement and background to contextualize why this research question is important. In the next two sections, I describe the data I obtained and the machine learning techniques I used to wrangle, visualize, and model the relevant data answer my research question. In the last two sections, I discuss the results of the analysis and my recommendations for future research.

**Background and Problem Statement**

The U.S. has the highest incarceration rate in the world, with over 2 million people behind bars. Research suggests that the incarcerated population faced disproportionate consequences of the pandemic due to poor conditions in these facilities such as high prisoner density, lack of room for social distancing, limited medical supplies, and inadequate sanitation, which were conducive to rapid community transmission of Covid-19. In the one-year period leading up to March 2021, these facilities recorded over 1,400 cases and 7 deaths per day on average.

State and local authorities undertook various policies to curb the spread of the disease among this vulnerable population including providing testing and protective equipment such as masks to incarcerated people and prison staff, reducing admissions and releasing prisoners, eliminating medical co-pays, and prioritizing the incarcerated population in early stages of the vaccine rollout. Policy responses in states, counties, and localities varied widely. For example, at the state level, the New Jersey legislature passed a bill that allowed for people with less than a year left on their sentences to be released up to eight months early starting October 19th. Similarly, at the county level, the Hays County sheriff announced a new “Cite and Divert” program in an effort to reduce arrests, jail time, and criminal charges.

In this project, I examine the extent to which the effects of the pandemic were concentrated among the incarcerated, by comparing their infection rates and death rates to those of their surrounding communities. I further explore whether there exists a relationship between regional economic and demographic characteristics and the performance of prisons or jails during the pandemic? Although several policies related to reducing the burden of the pandemic on the incarceration population were implemented at the state level, I choose to conduct my analysis at the county level so as to expand my dataset beyond 50 observations.

In order to develop the framework for my models, I reviewed literature investigating county level predictors of 1) Covid-19 infection rates in the general population and 2) incarceration rates. For example, XX found that XX was related to XX, and YY found that… Thus, for my model I choose economic variables as poverty rates, unemployment rates, and median household income, demographic variables such as race (percent of the population that is White) and education, and health indicators such as the percent of the population that is uninsured as my predictors of the performance gap in infection rates between the incarcerated population and the general population in the county.