Reflections

I found the collaborative activities in this assignment to be very helpful in understanding how to identify the effects of policies in time series data analysis. I collaborated with my friend Syed Obaid Dawarki, and we exchanged several ideas about the steps we should take to reach our final solution. It was also helpful to see how other people approached the same problem by checking the discord group for this class.

A variety of research methodologies, including case-control studies, time series, interrupted time series, case studies, and epidemy mathematical modeling, have been utilized to evaluate the effectiveness of Covid-19 prevention policies. In this assignment, we decided to go with the interrupted time series analysis method.

When governmental interventions are implemented at a specific time, they can cause an interruption in the trend of ongoing Covid-19 time series (e.g., new Covid-19 case time series, number of deaths caused by Covid-19 time series, etc.). Therefore, interrupted time series is a suitable statistical method to analyze the effect of governmental interventions. If changing the governmental policy caused an interruption in the pattern, the level and slope of time series and intervention parameters are statistically significant.

Based on the research questions posed in this assignment, the first thing that I learned from this research is that there is often a delay between the time of infection and the time a case is confirmed. This delay can be caused by several factors, including the disease's incubation period, the time it takes for symptoms to appear, and the lag time between infection and the availability of testing results. This delay can have important implications for the spread of the disease, so it is essential to consider it when modeling the pandemic.

The second thing I learned from the assignment is that masking may make it longer to get infected, or it may prevent some percentage of infection. We should consider a mask's effect when deciding whether to wear one.

The possibility of collaboration on this part helped me to think more deeply about the problem and to consider different perspectives. It also helped me to see the problem from a different angle and to understand the potential solutions better. Also, it allowed me to see how other

colleagues might think about the same issue and their perspectives. This helped me to come up with the steps to reach the final solution. The details for the same have been given below:

- I started with understanding the column names for all the three datasets
- I did the visual inspection of the data, and after that, I loaded the data using pandas in my Jupyter notebook and started the exploration of the data.
- With exploration, I learned that there is a lot of missing data in the data frame. I used the backfill and forward-fill methodology for the data with the infection numbers to

deal with the missing data. This methodology is generally used when working on time series data.

- I used the same methodology (backfill and forward fill) for the mask mandate data set.
- By merging the mask mandate and infection rates data frame, I started understanding the effect of the interruption in the time series data.
- I started with the OLS methodology, a very Naïve approach, and then used the ARIMA model for my analysis.
- ARIMA model helped me understand the mask mandate's effect on the rate of infections. The constrafactual graph showed a hypothetical view of how the infection would have progressed without a mask mandate policy.