Course: DATA 512 A – Human-Centered Data Science **Name:** Sravan Hande
Assignment: Project Part-1 Reflection Statement **Student Id:** 2129885

The collaborative model is beneficial in Data Science efforts and human-centered design activities since it plays a significant role in giving varied and constant feedback to enhance the overall design. Given the anonymity of the virus dynamics and the lack of knowledge, for this purpose. It is essential to provide and receive feedback in a collaborative cohort environment about the study's intricacies. The deep research and creation of the final visualization benefited greatly from discussing and testing several concept designs, techniques, and implementations to arrive at a robust solution.

Regarding the lessons learned from the assignment, the first important one is that all data collected and used for analysis must be adequately cleaned. The collaboration allowed me to understand the detrimental effects of these data gaps and outliers, and I was able to eliminate them from my county's statistics, where deaths are underreported on weekends and cases are reported as negative for a few days. A few subtleties may be missed because of ratification side effects, the professor explained in the week five lecture, depending on the scope and considerations used to gather the data. Given that the data is a time series, I cleaned the data by applying moving averages to smooth out the outliers on a few dates.

Another important finding showed that more than 73.4% of the population "always" wears masks, and 95% of the population does so frequently. This difficulty with hiding survey data collection exists. The 200 closest responses are used as the foundation for the estimation. Based on a small number of data points, I question the generalization of masking. The fact that they weigh the closest and farthest places even when the weights are not specified raises the possibility of inaccurate computations and misinterpretation if it is done in an arbitrary manner. Some of my classmates had the same problem, so we spoke about the benefits and drawbacks of having such a dataset. The impact of vaccination is my primary concern. The analysis's two graphics don't account for the vaccination state while making calculations. We learned that we may use the SIRV model with a valid vaccination rate (p) and reinfection rate to simulate the effect of vaccination on the status of the disease (alpha). The calculations for infection rates are somewhat altered as a result, but the plot remains largely unchanged.

We all concurred during the discussion that the impact is not always immediate. After a few time intervals, the shift may be noticed in the data, and we tried to determine how it differs for other counties with strict/no masking laws. Even though the initial phase of vaccination was initiated, and most counties had a masking mandate throughout the month of December 2020, there was a rise in instances. The reason might be that during the holidays of Christmas and New Year, a lot of people travel, making masks less effective at stopping the spread of sickness. we can observe that after the initial vaccination took place in December and the holiday season had passed, the infection rate gradually began to decline.

In the beginning, I experimented with different derivative analytic methods to comprehend the shift in virus spread. However, Charles Reinertson's assistance allowed me to familiarize myself with the idea of change point detection. After doing some investigation, I came across a program called ruptures that uses regression splines to fit the data and pinpoint the critical locations where the derivative/slope changes. After discussing how these change points could be translated to the mask mandate policy, I then shared it with the cohort because I thought it was highly comprehensible. Along with Hriday Bhagar, I also spoke on the SIR modeling, where we attempted to incorporate the reinfection rate and recovery status into our model. It is extremely fruitful how this collaborative activity brought together individuals with similar backgrounds and research interests while also allowing them to share various schools of thinking to develop a solid solution.

On the other hand, I was able to understand several change point detection choices thanks to my lack of expertise with time series tools. Eli Copron's Pelt Search technique and Charles Reinertson's Prophet tool, respectively, were used. To plot the graphs, I modified some of Eli's business logic. A portion of the data analysis benefited from discussing several time series that were tied to different terminology. This collaboration limited my thinking in one area because there might be simpler ways to comprehend changes in time series without using complicated change-point detection techniques. It is usually preferable to use more straightforward techniques when it comes to the explainability of your data science solutions.

In conclusion, working together and brainstorming with others greatly aided my ability to generate fresh concepts, confirm my presumptions, and obtain support for tools and strategies as I overcame obstacles in the task. When there were numerous ideas on the table and it was difficult to weigh the advantages and disadvantages of each to arrive at the best deterministic solution, this might often impede development. I was able to express my ideas more freely and use a couple methods that I found intriguing. This freedom in collaborating allowed me to complete my research more quickly and concentrate more on comprehending and implementing the specifics of the study into my model.