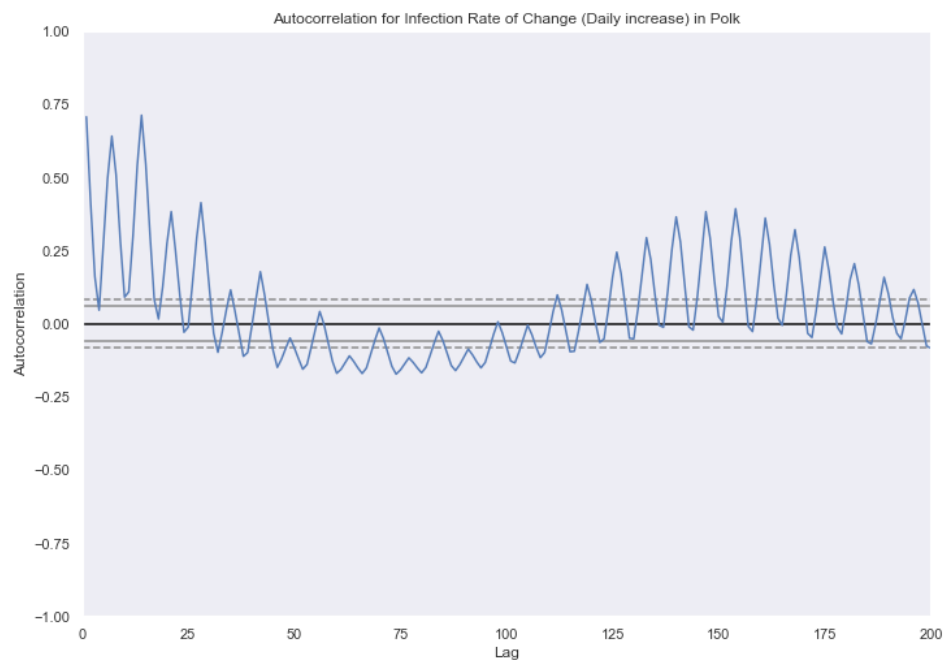


Project Part 1 Writeup

Visualization Explanation

The figure below is created to show what could be the influencing factors of the daily increase of Covid 19 cases in Polk, Florida between 2020-1-22 and 2022-11-1. This autocorrelation shows the correlation of the daily increase data with itself with a time lag. The x-axis represents the size of the lag in terms of days. For example, a 25-day lag compares the data[0:n-25] with data[25:n], indicating whether there could potentially be an impact of the past data on the future. The y-axis, on the other hand, represents the strength of the correlation. The closer the values are to 1 and -1, the stronger the correlation, and the closer the values are to 0, the weaker the correlation is. The horizontal lines represent the 95% confidence interval and the dotted line shows the 99% confidence interval.



From the figure above, we could notice that for lag 1 to 25 days, there is a strong evidence that there is some level of autocorrelation, as well as lag of 125 to 200 days. This implies that the daily increase of the past 1-25 days, as well as the previous 125 to 200 days are highly correlated to the daily increase of the present. However, the correlation does not imply causation, but this could be something we could further investigate in the next steps of the project.

Reflection of Collaborative Activities

Mostly the collaborative activities occurred on the Discord group chat, where I obtained lots of insights from others. Arik provided some of his code for data processing, which accelerated my data understanding process. Charles, Tharun and Eli all gave some examples on how we could do change point detections on the time series data, using tools including Facebook Prophet, Ruptures and Pelt Search Method, which were novel concepts to me.

Also, I found people whose assigned state doesn't have any mandated mask policies like me. At the beginning, I was confused on how to model the impact of masks without any mask policies information given. I posted my confusion and luckily I got feedback from Amrit and Urmika, pointing out that the daily increase (derivative) data themselves could be a propelling factor for people to put their masks on, and that gave me the idea of creating an autocorrelation plot for the daily increase rate. I also got inspired on cross correlation plots for daily increase rates for Polk, and other time series data such as daily increase for entire US, the total confirmed cases and death cases for US and Polk. Those could all be motivating factors influencing the voluntary mask rates.