Review of
“Exploration of COVID-19 and the Global Climate” (group 46)

Score: 2.5

Summary: In this project the authors tried to address several unrelated questions. The first question is about general clustering of the changes in temperature by region, elevation etc. Some preliminary EDA in this direction has been performed, but there are no results presented for this question. The second question is dedicated to exploration of the effect of party affiliation on the spread of covid-19. The presented result is that at the onset of the pandemic Democratic states had more cases but later the trend reversed itself and now Republican states have more cases. Finally, the last question is whether there is correlation between COVID-19 and climate. Here the authors present two scatter plots of mortality rates vs average temperature that show no clear correlation.

Rationale for the score: The fact that this project is a collection of unrelated questions with only superficial analysis of each instead of an in-depth analysis of one question lowers its value. While all three questions raised by the authors are interesting each of them deserves a more thorough investigation. In particular, all methods employed in the project are very basic and could not discern dependencies/correlations in most cases. For the first question, the problem is not clearly formulated, the authors only showed their EDA and the methods that they are planning to use are not presented. For the second question, while a result (a plot of cases for red/blue states, interestingly enough red is replaced with orange) is presented and authors' guesses for its origin are given it does not rise to the level of causal inference since no testing of hypothesis is performed. As for the last question, a very superficial analysis has been performed and it did not yield any valuable result. The model is rife with confounding variables and other parameters that should have been controlled for, such as population density, presence of cases in the past (the temperature should affect the rate of spread not the total number of cases), government measures, build up of herd immunity, etc.

I do not have a rubric for projects that consist of many unrelated parts. An ideal project with a single question (corresponding to the score of 5) should address an interesting question, use appropriate methods to answer it, implement these methods adequately and use additional evaluation to show that modelling is valid.

3 strong points:

1) While the results are not very strong, the visualization very well represented those results
2) In the cases when the analysis did not yield the desired result the authors seem to have correctly captured shortcomings of their analysis
3) In the case of the effect of temperature on the spread of COVID-19 a comparison with previous research on the topic is done and a possible explanation of the discrepancies is given

3 weak points. Apart from the scattered focus and unclear result (or even ways to arrive at them) the weak points are

1) In the climate-COVID-19 study causal inference has not been done properly, while a simple scatter plot does not show correlation, controlling fro other variables might have improved it
2) In the party-COVID-19 study a claim about causal dependence is made but is not verified in any way (the project should be about causal inference)
3) It is not clear what cases per 100k mean on the party/COVID-19 graph. Is it average over the rates of all states or is it per the total population of all blue/red states?
4) The first question lack any description of what method will be used for clustering (such as K-means)

Suggestions:

1) Pick one question and focus on it, or try to connect some of the questions with each other
2) Add some causal hypothesis to some of the questions and test it with the method of causal inference. Such as does the party affect COVID-19 spread? In order for this to work some variables should be controlled for and the model should be evaluated
3) In the case with temperature/cases, doing a time series analysis seems more appropriate, as the temperature should control the rate of spread not the total cases
4) Same goes for party affiliating, it should control how COVID-19 spread not the total cases at the moment