Project 1

Due February 18th, 2024

2:00 PM 100 points

CS 4432/5599

MATH 4463/5563

Data Science and Applied Machine Learning

1. Using your team's knowledge of pandas, NumPy, MatplotLib, etc. to perform an analytical study on the United_States'_COVID-19,_Community_Levels_by_County dataset posted to Moodle. Prepare a 10-minute presentation based on your analysis. You are welcome to use any additional data/resources, but make sure to discuss them in the presentation and include them in your submission. The scope and focus of the presentation are up to you as a team, however, keep the following guidelines in mind:
    a. All team members must be part of the presentation, any team members who do not, will not receive credit.
    b. Your presentation should state the goals of your research, i.e. what did you wish to learn through your analysis.
    c. Your presentation must include at least one of the following visualizations:
        i. Pie Chart
        ii. Histogram
        iii. Bar Chart
    d. Teams are encouraged to explore and utilize other visualizations based on the nature of their findings .
    e. Presentations should include the following sections:
        i. Introduction/Background
        ii. Goals of Research (What specific insights or questions do you aim to address through your analysis?)
        iii. Methodology (i.e. what analysis did you preform)
        iv. Results
        v. Analysis of the Results
        vi. Conclusions/Reflection (i.e. what did you learn from the analysis)
2. Submit a video with your presentation, any material you used during the presentation, such as PowerPoint slides, any code that was used to perform your study (well-documented and commented) and if necessary, any other data/resources used (discussed in the presentation).

Notes:

1. All four professors will score your presentation out of 25 points, your team's final grade will be the sum of those scores.
2. Your grade will factor the quality of your presentation, analysis, and visualizations, as well as the depth of your analysis.
3. After all submissions have been graded, your presentations will be posted for all students to view on Moodle.
4. The data provided is from the United States Center of Disease Control and can be treated as a reliable source.
5. This project will account for 10% of the final grade in this class.
6. All data analysis/visualization must be done using Python.
7. Teams have been assigned for this project (see attached file).