To do:

1. Perform exploratory data analysis (EDA)
   1. at least two data visualizations
2. Describe any data cleaning or transformations that you perform and why they are motivated by your EDA.
3. Apply relevant inference or prediction methods
   1. (e.g., linear regression, logistic regression, or classification and regression trees),
   2. including, if appropriate, feature engineering and regularization.
4. Use cross-validation or test data as appropriate for model selection and evaluation.
   1. Make sure to carefully describe the methods you are using and
   2. why they are appropriate for the question to be answered.
5. Summarize and interpret your results (including visualization).
6. Provide an evaluation of your approach and discuss any limitations of the methods you used.
7. Describe any surprising discoveries that you made and future work.

Format: Gradescope as a zip

1. Jupyter Notebook. A single notebook that contains all the code run for the project.
   1. The code should be clear, easy-to-read and well-documented.
   2. explain in text what you can conclude from the output of your code cells.
2. Project narrative.
   1. This typed PDF document should summarize your workflow and what you have learned.
      1. a title, list authors, abstract,
      2. introduction, description of data, description of methods,
      3. summary of results, and discussion.
      4. Make sure to number figures and tables and include informative captions.
   2. Specifically, you should address the following in the narrative.
      1. page limit of 6 pages, excluding figures and tables.
      2. What type of question are you trying to answer with the data
      3. Perform exploratory data analysis (EDA) and provide data visualizations.
      4. Describe any data cleaning or transformations that you perform and why they are motivated by your EDA.
      5. Carefully describe the methods you are using and why they are appropriate for the question to be answered.
      6. Create a complete statement of the model and assumptions they are using for inference
      7. Summarize and interpret your results (including visualization).
      8. Address the following 7 specific questions. –
         1. What were two or three of the most interesting features you came across for your particular question?
         2. Describe one feature you thought would be useful, but turned out to be ineffective.
         3. What challenges did you find with your data? Where did you get stuck?
         4. What are some limitations of the analysis that you did? What assumptions did you make that could prove to be incorrect?
         5. What ethical dilemmas did you face with this data?
         6. What additional data, if available, would strengthen your analysis, or allow you test some other hypothesises?
         7. What ethical concerns might you encounter in studying this problem? How might you address those concerns?
      9. Provide an evaluation of your approach and discuss any limitations of the methods you used.
      10. Describe any surprising discoveries that you made and future work.

Covid-19 The coronavirus datasets contain different information about COVID-19. They include information about daily cases and fatalities, as well as hospital level data. We provided multiple datasets for this topic and you are required to use at least one of them if you choose to work on this topic. The sources of the datasets come from:

<https://github.com/Yu-Group/covid19-severity-prediction/tree/master/data>

<https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data>

The links of the four .csv files provided for this topic are in the README file under the covid19 data folder. Below are some potential directions for this dataset:

1. What factors can be beneficial in predicting how fast coronavirus spreads in a region?
2. Suppose you wanted to model the number of coronavirus cases in a region.
3. What kind of model and features would you use?
4. How might historical trends inform our predictions on the number of coronavirus cases in a given day?