

Assignment 5: Data Analytics (Spring 2021) (4000 Level - 5%, 6000 Level – 10%) Oral
Due: Monday, March 15th, 2021 (Presentation slides due at the beginning of the class)
Presentation method: virtual presentation during the class time on March 15th and March 18th

Slides MUST be submitted to LMS at the beginning of the class on **March 15th, 2021**.

Late submission policy: if you do not present virtually by Friday, 19th March 2021, you receive 0% for this assignment and your project is not approved (and you still need to schedule this presentation). This may delay your start on your project and affect its completion.

Note: Your presentation for this assignment should be the result of your own individual work. Take care to avoid plagiarism (“copying”), and include references to all web resources, texts, and class presentations. You may discuss the project with other students, but do not take written notes during these discussions, and do not share your presentations before class.

General assignment: Term project proposal. This is a chance to get feedback on what project (questions, data, methods, etc.) you currently plan for your term project. The grade is NOT based on presentation quality. Grade is based on covering content detailed in #1a-f below.

Guidance: Your term projects should fall within the scope of a data analytics problem of the type you have worked with in class/ labs or know of yourself – the bigger the data the better.

This means that the work must go beyond just making lots of figures. You should develop the project to indicate you are thinking of and exploring the relationships and distributions within your data. Start with a hypothesis, think of a way to model and use the hypothesis, find or collect the necessary data, and do both preliminary analysis, detailed modeling and summary (interpretation). Details of what will be required will be in Assignment 6.

Note: What you present in this assignment does NOT have to be what you eventually conduct your project on. This is to get you to start thinking about how an end-to-end project would look. You can propose to use datasets made available by data-challenge competitions, but you CANNOT use their hypotheses or challenge questions. You must develop your own hypotheses/questions.

4000 Level (5%)

Note: There may be methods you use that we have not covered – that is okay.

1. Oral presentation (5 mins). Suggest these slides (~ 45 sec. each after title slide):

- a). Title (with your name)
- b). Problem area – why it is of interest (in general or to you), what you might want to predict? This could be a hypothesis.
- c). The data – where it might come from, why it may be applicable, and any preliminary assessment you’ve made?
- d). How you plan to conduct your analysis: distribution, pattern/ relationship and model construction. What techniques do you think you will use?
- e). How do you plan to apply the model? What are the possible uncertainties?
- f). What do you want to predict and what decisions (prescriptions) may be possible? What would a good outcome be?

6000 Level (10%)

Note: There may be methods you use that we have not covered – that is okay.

1. Oral presentation (6 mins). Suggest these slides (~ 1 minute per each after title slide):

- a). Title (with your name, Level 6000)
- b). Problem area – why it is of interest (in general or to you), what you might want to predict? This could be a hypothesis.
- c). The data – 6000 Level students must have minimum of 2 datasets* for the project. where it might come from, why it may be applicable, and any preliminary assessment you've made?
- d) How you merged (combined) two or more datasets (how did you do the data synthesis on two or more dataset)
- e). How you plan to conduct your analysis: distribution, pattern/ relationship and model construction. What techniques do you think you will use?
- f). How do you plan to apply the model? What are the possible uncertainties?
- g). What do you want to predict and what decisions (prescriptions) may be possible? What would a good outcome be?

NOTE*: 6000 Level Students must have a minimum of two datasets (or more datasets) and must use them to combine and create the project dataset that you are planning to conduct the analysis. Two datasets should have a meaningful relationship when you combine them. If you are working on NASA IMPACT or any other NASA datasets that provided by the instructor, it is sufficient to work on a one dataset: Example: If you are planning to work NASA IMPACT smoke dataset, that qualifies as your dataset for the project. Please come and talk to the instructor if you have any questions or doubts during the One-on-One sessions or during office hours.