

Assignment 3: Data Analytics (Spring 2026) Project Proposal - Oral (5%)

Due: February 24th, 2026

Presentation method: Presentations during the class time on February 24th and 27th.

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.

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. The grade is based on covering content detailed in (1) a-c below.

General guidance for the project: 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. You should develop the project around a research question or interest that you have. You will locate and acquire a dataset that you believe contains knowledge relevant to your research question. Typically, you will start by exploring the relationships and distributions within the data. This would be followed by exploratory analysis to reveal structure and patterns in the data, relationships between variables, etc. Once a specific hypothesis is formed, predictive models would be developed, evaluated and interpreted before final conclusions can be drawn.

Details of project activities and requirements will be released in **Assignment 4**.

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 what planning for an end-to-end analytics project would look like.

You may propose to use datasets made available by data-challenge competitions, but you must **NOT** use their hypotheses or challenge questions. You must develop your own hypotheses/questions.

4000 Level

1. Oral presentation (5 mins).

a). Problem area: why it is of interest (in general or to you), what might you want to predict? This could be a hypothesis.

b). The data: what is the source of the data? why the data may be applicable, and any preliminary assessment you've made (e.g. size of dataset, data quality, etc.).

c). The analysis plan: how you plan to conduct your analysis relative to exploring distributions, detecting patterns/relationships, and developing predictive models. What are model inputs/outputs? What techniques do you think you will use? What would a good outcome be?

6000 Level

1. Oral presentation (5 mins).

a). Problem area: why it is of interest (in general or to you), what might you want to predict? This could be a hypothesis.

b). The data: what is the source of the data? why the data may be applicable, and any preliminary assessment you've made (e.g. size of dataset, data quality, etc.). *This should cover the multiple datasets. Additionally, what is the relation between the datasets and why is that useful for the analysis?*

c). The analysis plan: how you plan to conduct your analysis relative to exploring distributions, detecting patterns/relationships, and developing predictive models. What are model inputs/outputs? What techniques do you think you will use? What would a good outcome be?

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

NOTE: 6000 Level students must have a minimum of two (or more) datasets and must use them both in the same analysis. The datasets should have a meaningful relationship between them. Please come and talk to the instructor if you have any questions or doubts, during the lab sessions or office hours.