STA 478 Project Expectations

Dr. Robert Buscaglia

November 03, 2025

Projects

Topics should be chosen from the website https://www.kaggle.com or https://ourworldindata.org/. The project will cover an analysis of the data set chosen. This will include evaluation of all variables with visualization and discussion, discussion of possible modeling techniques, application of methods from within the course to the data set, clear and precise presentations of results. The projects will be worked on individually.

A written proposal will be completed first. The projects will culminate in written essays & in-person presentations. Information for each section is presented below. The project will constitute 25% of the overall grade and will count as our final exam. It is imperative you work diligently on these projects and present as much in-class knowledge, code, and conceptual understanding as possible. This includes also using statistical language and testing covered in pre-requisite courses.

All students will present 5-minute flash-talks during our scheduled final exam. Details will be discussed in-class and all requirements will be made clear in announcements. This portion will be combined with the written essay.

Written Proposal [20 points]

The written proposal will be due *Friday, November 21st, 2025 before 11:59 PM*. Please prepare a written proposal following the outline below.

- Must be written in RMD, turned in as a PDF
  - You may use options to alter the document.
  - It must be clear and easy to read.
- 2-page written limit
  - Not including figures/tables.
- Data set and reference
  - Provide a link to the online resource where the data was downloaded.
  - Discuss any troubles you (may) have in cleaning the data.
- Introduction
  - Briefly introduce the topic to be studied.
  - Why did you choose the topic?
  - Discuss the data thoroughly.
    - Use tables and figures to make it clear what variables, types, and amount of data you have.
    - Use this as an opportunity to do data cleaning.
    - The proposal main goal is to show you have a data set worth analyzing!
- Analysis
  - Consider what models you think will work.
    - Which models might you like to incorporate?
    - What model types might you like to learn more about?
  - The models used must be within the scope of the course and what has been taught.
  - You will be required to build one inferential model and one predictive model, at minimum.
- Goals
  - Discuss what you hope to complete by the end of the semester.
  - Set goals for your project completion, this is 1/4 of your grade!
  - These should be the main accomplishments you see from doing the project.
    - Analysis goals?
    - Project result goals?
    - Personal goals?

Written Essay [40 points] & In-Class Presentations [40 points]

The written essay will be due *Tuesday, December 9th, 2021 before 11:59 PM.* Please prepare a written essay following the outline below.

- Must be written in RMD, turned in as a PDF
- 4-page written limit
  - Not including figures/tables
- Figures and Tables should each be a separate page
  - Each figure/table should also include a caption/discussion
  - No figure/table should be shown that isn't discussed within the 'written' pages.
- Sections:
  - Introduction
    - Data and topic / Reason behind analysis (See proposal)
  - Data evaluation (Materials)
    - Variables and descriptions
    - Images and Tables for presenting data.
    - Review of data cleaning
  - Modeling introduction (Methods)
    - What was chosen and why?
    - How were the models implemented?
    - Are there any limitations?
    - At least one inferential model must be analyzed
    - At least one predictive model must be analyzed
  - Analysis results
    - Cross-validation must be performed
    - Present a comprehensive set of findings regarding your results
    - Use tables and figures to present results clearly
  - Discussion of final models and analysis.
    - Present final models and discussion
    - Interpretations of the inferential model MUST be included
    - Interpretations of the predictive model are optional and can be included if possible.
    - Prepare summary information for inferential models
      - Include confidence intervals and proper interpretations.
    - Compare and contrast two or more predictive models if desired.
  - Conclusions
    - Wrap things up nicely
    - What have you learned about the project topic?
    - What have you learned about the modeling aspects of the project?
    - Pros/cons - what would you change or try in the future?
  - Citations
  - Appendix
- Code must be annotated and included in appendix.
- Data must be submitted as a .RData file in a cleaned and ready-to-use format.
  - Data needs to be provided to me (this is why we are using available data).

You will wrap the "highlights" of this work up into a 5-minute in-class presentation to be done during our pre-set final exam time of Tuesday, December 9th, 2025 from 12:30 – 2:30 PM. Additional details on this portion of the work will be given during class and as announcements.