

STAT535 Final Project

Due Date: Wednesday, December 15, 2021, 6:00pm EST

Your final project involves building and analyzing **datasets from internet** and using the datasets to **make your prediction** on related topics. You are allowed to discuss and share ideas with others but the actual work should be done by yourself or by your group.

Overview

The final project is a 12 minutes group presentation with up to 5 more minutes for questions and comments. Each group will consist of three students. Each member should be in charge of and present at least a part of the project even though everybody in a group should work together to make a complete presentation. Everybody in a group will earn the same grade, so you should help each other. **The time limit will be strictly enforced.** You MUST rehearse your group presentation several times before presenting to the entire class.

Process

1. Before **11/22/2021, 11:59pm**: form groups of three and notify me of those. Students who will not notify me of their group affiliation by this date will be randomly assigned to groups.
2. Discuss with me via a Zoom meeting your tentative ideas for a project.
3. **Submit an abstract by Wednesday 12/8/2021, 9:00pm** (last day of classes). The abstract should include:
 - Title of project
 - Full names of group members
 - A brief description of the project and data collection (website, date type etc.)
 - Objectives and interests (what will be predicted)
 - Class material to be used for the analysis
 - A description of the responsibilities of each of the group members.
 - Conclusion (If you don't have the conclusion yet, please just present it on the presentation day.)
4. **Presentation 12/15/2021, 6:00-8:00pm in class** (this is the time designated for the final exam in the class). The order of the presentations will be randomly assigned.

Data Collection

Use web-scraping covered in the course. Use R.

Visualization of Datasets, Analysis

Use visualization functions in R for presentation purpose. Use course material (packages, functions, numerical/statistical schemes for optimization, sampling schemes, fittings etc.) as needed. You must include in your analysis at least one Monte Carlo method or a Bootstrap technique.

Presentation

Use Rmarkdown with the presentation output mode.

Grading Criteria

1. Outline: Abstract turned in on time? Is it well written? Demonstrates a good understanding and interpretation? Includes all the items as in the instruction? (10 pts.)
2. Slides: Legible? Well-organized? Labels for graphs? Clear structure? Good visualization? Good representation? (10 pts.)
3. Presentation: Well organized? Good communication? Fairly distributed to each member? Cooperative? Comprehensible? Transition between speakers? Speed? Interaction with audience? (10 pts.)
4. Content quality: Large sample size? Creative? Understanding material? Correct calculations? Assumptions properly highlighted? Proper analysis and method? Logical flows and conclusion? Reasonable prediction? (10 pts.)
5. Time: Fairly distributed to each member? Too long (more than 10 mins)? Too short (less than 7 mins)? (5 pts.)
6. In addition to presenting your own work, you will be asked to submit your comments about the presentations of the other groups in writing. These comments will be shared with the group presenting (anonymously). The last component in your grade then will be my evaluation of the validity and quality of your feedback to others (5 pts.)