

Final Project Guidelines

Modern Data Mining

(Due: Wednesday, 12/12, before 11:59 PM)

Instead of doing a final exam, we ask you to do a project. An important aspect of learning is through hands on experience, especially when working with data and analysis – nothing can prepare you better for the real world than analyzing a real data set which you're interested in.

We hope that you find the project rewarding and worthwhile. This final project is a good addition to your resume or CV. They are very attractive, valuable sets of skills/experiences!

You should demonstrate the following skills:

- Statistics
- Domain knowledge
- Software proficiency
- Written and verbal communication skills

Prompt: Carry out a study applying the methods that you are learning this semester.

Requirements:

- 1) An interesting topic, preferably with application value, and originality (preferred)
- 2) Data set: submit both the *original data*, and a *cleaned/processed version*.
- 3) Techniques: materials covered this semester, or explain further if they are beyond this course
- 4) A complete report: As a culmination of your work this semester, it is very important that you control the R output to keep the report readable. All directly relevant R output should be clearly labeled in an appendix, but should be found easily (e.g. via footnotes). The following is a suggested outline:
 - a) Executive summary
 - b) Goal of the study
 - c) Data
 - d) Findings
 - e) Detailed analysis
 - f) A reproducible R code (with comments)
 - g) **Maximum of 15 pages** for the main body of the report

5) Proposal

We will not collect your proposal.

6) Due date: Wednesday, 12/12, before 11:59 PM to Canvas

Other notes:

- a) This is a group project.
- b) **You must submit your documents to Canvas with a label of your last name and a key word about the topic, e.g. “Zhao_ElectionPrediction”**
- c) **The submitted documents should contain:**
 - **Final report**
 - **.rmd users: your .rmd file, and compiled pdf or html. Must submit compiled file!**
 - **word users: your word file, the pdf version, and R code**
 - **All the original data and the clean data used**