Stat 628: Data Science Practicum

Module 2 Grading Guidelines

Deliverables and Deadlines:

|  |  |
| --- | --- |
| Deliverables (for **Tuesday Group**) | Due Date (All times are on Central Time) |
| Presentation slides (.pptx, .ppt, .pdf)  Note: You’ll give an in-class presentation on Tues. Oct. 18. | Monday, Oct. 17, 2022, 11:59pm |
| Two-page executive summary (.pdf) | Monday, Oct. 17, 2022, 11:59pm |
| Github repo containing code (a link) | Monday, Oct. 17, 2022, 11:59pm |
| Shiny (or Web-based) app (a link) | Monday, Oct. 17, 2022, 11:59pm |

|  |  |
| --- | --- |
| Deliverables (for **Thursday Group**) | Due Date (All times are on Central Time) |
| Presentation slides (.pptx, .ppt, .pdf)  Note: You’ll give an in-class presentation on Thur. Oct. 20. | Wednesday, Oct. 19, 2022, 11:59pm CST |
| Two-page executive summary (.pdf) | Wednesday, Oct. 19, 2022, 11:59pm CST |
| Github repo containing code (a link) | Wednesday, Oct. 19, 2022, 11:59pm CST |
| Shiny (or Web-based) app (a link) | Wednesday, Oct. 19, 2022, 11:59pm CST |

**All deliverables** **will be submitted to Canvas**. For the presentation and two-page summary, you must submit (i) one file for the presentation and (ii) one file for the two-page executive summary, all on Canvas. For the Github and Shiny app, you must provide (i) a link to the Github page that is hosting all the code and (ii) a link hosting your functioning Shiny App in a Shiny App server of your choosing (or an equivalent web-based app), all on Canvas. If possible, we encourage you to submit all the deliverables only once.

It is **your responsibility** to make sure that all the deliverables are submitted on time for your lecture group (Tuesday or Thursday). No late submissions will be accepted. We'll only grade the latest submission that were submitted before the due date.

Each deliverable is graded on a 1-3 scale.

Groups:

You will work in groups of three or four. Each group will be responsible for all the deliverables. Groups will be randomly assigned and your group assignment is available on Canvas.

Presentations:

The goal of the presentation is to practice presenting your statistical findings in a concise and clear manner. The presentation should include key evidence (e.g. plots, tables, inferential methods, etc.) that support your findings. Your presentation must be clear and precise enough that **any employee with an quantitative background** (not necessarily in statistics) should be able to understand what statistical analysis you used and how you have reached your conclusion. The exact grading rubric for the presentation is outlined below. If you would like, you can also demo your Shiny app (see below), but this is not a requirement.

Your group will prepare a **5 minute** in-class presentation of your data analysis; your presentation can be shorter than 5 minutes, but it cannot be longer than 5 minutes. All members of the group must work on the presentation and speak **for at least one minute** during the in-class presentation. The exact time of your group’s presentation will be determined randomly on the first day of the presentation.

Due to time constraints, the time limit will be**strictly enforced**. To encourage this behavior, every additional 30 seconds after the time limit will incur a penalty of 0.5 points. It is **your responsibility** to rehearse your presentation so that it stays under **5 minutes.**

Each group will submit **a single** presentation slide (in .pptx, .ppt, .pdf) to Canvas.It is your responsibility to check with me or the TA that your slides can be displayed properly on the projector in the lecture hall before the presentation day.

Two-Page Executive Summary

Your group must submit a **two-page** executive summary of the data analysis, with at most **one additional page** for references & contributions. That is, all the statistical analysis, plots, tables, figures must fit in the two- page limit and the one-page reference section **can only be** **used** for references to other works & contributions (see below).

Your summary must include (i) your overall findings and the rule of thumb, (ii) relevant and important evidence to support your findings (e.g. plots, tables), and (iii) important details of your statistical analysis (e.g. type of model used, inferential quantities, outliers, leverage points, modeling assumptions, etc.). Your summary should be detailed enough that **any data scientist** can **understand** your summary and **replicate** key statistical analysis. The exact grading rubric is outlined below.

On the reference page, you may follow any reasonable style for references (e.g. MLA, APA, Chicago Manual of Style, etc.).

All members of the group must contribute to the summary. Specifically, on the reference page, the group must clearly indicate **each member’s contribution** **to the summary, the presentation, and the code**. For example, you can say that

1. HK (initials of your group member’s name) wrote/edited the diagnostic part of the summary, worked on slides (blank) to (blank). HK also created code related to data cleaning , revised/ maintained the code related to Figure 3, and is ultimately responsible for the data cleaning portion of the code.
2. BK (initials of your group member’s name) wrote/edited the model building and interpretation part of the summary, worked on slides (blank) to (blank). MK also created/edited/maintained code to model building (e.g. lm(), summary(lm()), edited code related to plotting, and is ultimately responsible for the model building portion of the code.

All of the summary must be typed in 12-inch Times New Roman or Sans Serif font, single-spaced, with 1-inch margins and (again) must include all relevant figures/tables, and equations. Note that all figures and tables must be legible when printed on a standard 8x11 paper.

Each group will submit **a single electronic copy (in .pdf)** to Canvas. It is your responsibility to submit the file on time and that it can be opened in a standard PDF or Word viewer.

Github Repository

Your group must publish a Github repository and provide **a web link** to the repository for grading. The Github repo must contain the following:

1. a data folder containing the raw and (if relevant) cleaned data
2. a code folder containing all the code for your analysis (e.g. cleaning the data, running the analysis, producing figures/tables, Shiny app base code, etc.)
3. an image folder containing any figures/images/tables produced in your analysis.
4. The two-page pdf summary file above.
5. **a README Markdown file** summarizing the contents of the repository and directions on how to use the code.

The code must **replicate every part of your analysis** from start (i.e. reading in the data) to finish (the figures/tables/results in your presentation and two-page summary). This includes, but is not limited to: data cleaning, outlier removal, model building, evaluation of different models, statistical testing, prediction, and any and all intermediary plots, tables, and analysis. Your R code must reproduce the **exact tables, plots, and other analysis** in your summary and the presentation (i.e. exact labels for axis, color shading plots, etc.). Also, your R code also must be **well-documented** so that **any undergraduate statistics or data science** student can read and understand your code. This is important for reproducibility and to track down potential bugs/errors in your analysis.

All members of the group must contribute to the codebase. It is **strongly encouraged that you use Github pull/push/commit** functions to manage your project files and to make it easy to record who contributed/worked on different parts of the code. Using Github this way (e.g. with push/pull/comits) will be required for Module 3.

Finally, we remind all students that it is **your responsibility** to make sure that the code is not copied/plagiarized/fabricated from unauthorized sources and (again) your code produces the results reported in your other deliverables (see reminder below).

Shiny App:

Often, data science jobs expect you to make “actionable” prototypes/products based on your data analysis. To this end, you will create a Shiny (or a web-based) application that will run your body fat calculator in real-time and **submit a link to your** **live/running Shiny app** for grading.

Shiny is an easy-to-use platform to turn your R analysis into web-based applications. While you do not have to specifically use Shiny (if you have app development experience, feel free to use any language/platform!), all applications must run on the latest Chrome browser. For more information about Shiny, visit: <https://shiny.rstudio.com/>.

We’ll leave the user-interface and other graphical specifications of the Shiny app up to you. However, your application will be graded on

1. whether it runs in real-time,
2. whether it is robust to erroneous inputs,
3. whether it provides useful and insightful information to the end user, and
4. whether there is some form of a contact information if the end-user has questions about the application. In particular, when publishing the Shiny App, the App must also contain contact information about who maintains the App so that bug requests can be directed to him/her/them.

Grading Rubric:

We will use the following grading rubric to grade your deliverables. Each deliverable will be graded on a 1-3 scale.

|  |
| --- |
| Presentation |
| 1. Clear, takeaway message (e.g. a “rule-of-thumb”) that is easy to use and accurate 2. Relevant, concise, and clear summary of statistical analysis 3. Relevant (**no extraneous and/or “R-dump” plots**!) and visually accurate plots 4. Strengths and weaknesses of the analysis 5. A simple, illustrative example to demonstrate the rule-of-thumb. 6. Overall, did the group present convincing evidence for their finding? 7. Overall, was the delivery clear and easy to understand? 8. Was the presentation under the prescribed time mentioned above? |
| Two-Page Executive Summary |
| 1. Introduction with clear motivation and thesis statement/rule-of-thumb. 2. Background information about the data 3. Motivation for the model used and statement of the model 4. Concise and relevant summary about estimation and inference of relevant parameters, which may include estimated coefficients, R^2, standard errors, confidence intervals, p-values, hypothesis testing statements, and etc. **No “data/R printout dump”** (properly format your tables/plots so they look presentable!) 5. Clear, laymen’s interpretation of the estimates and inferential quantities 6. Model diagnostics and checking modeling assumptions with plots 7. Strengths and weakness of the group’s data analysis 8. Conclusion 9. Does it follow instructions concerning the two-page limit? |
| Github Repository |
| 1. The Readme file is concise and summarizes the contents of the repository 2. Contains clean, readable, well-documented, and error-free code 3. Data can be easily read and cleaned using the code provided 4. Figures/tables are legible, concise, and clear 5. Contains the two-page summary file. |
| Shiny Application |
| 1. Does it run in real time? 2. Is the application robust to user inputs? 3. Does it provide useful and insightful information to the user? 4. Is there contact information in case if users have any questions about the app? |

A Reminder: Academic Integrity

Each year in Stat 628, we have a few students who violate the academic integrity standards laid out in class. We take this opportunity to remind students of the policies regarding academic integrity.

Each student assumes the responsibilities of an active participant in UW-Madison’s community of scholars in which everyone’s academic work and behavior are held to the highest academic integrity standards. Academic misconduct compromises the integrity of the university. Cheating, fabrication, plagiarism, sabotaging other groups’ work, unauthorized collaboration, and helping others commit these acts are examples of academic misconduct. Specific examples include, but are not limited to,

1. Copying, plagiarizing, stealing, fabricating any of the deliverables, especially the code or the plots/tables, from other groups, students outside of the class, or the web. In particular, while you may ask other groups for general ideas and questions, you cannot ask for help cleaning the data set, analyzing the dataset, and doing other activities that would be inconsistent with the academic integrity at UW-Madison. If you are unsure, you are always welcome to ask the TA or the professor.
2. Using unauthorized sources, especially using someone else’s analysis of the dataset on the web, either in its entirety or in parts. You are also not allowed to copy, steal, plagiarize, paraphrase, or use any analysis that was already conducted on this data (or a derivative thereof) from others (e.g. data science courses online, someone’s blog post or R markdown, Google Cloud’s API platform, AWS Machine Learning library examples, or Azure ML, any pre-written software/code specific to this data or derivative thereof, etc.).

However, you are **strongly encouraged** to browse through resources on body fat, health, and other relevant information to gather **background information**. You are strongly encouraged to use the information from your background research **to complement** your own analysis and **provide proper attributions**. In short, your analysis of the data must be **original** and **must be your own work**. Or, in industry-lingo, you should not be stealing others’ intellectual property.

If you have any questions about this, please come talk to the TA or the professor.

1. Sabotaging others’ work by deleting, copying, damaging, misrepresenting, or falsifying information about the data, any of the deliverables, or the project.
2. You may not ask someone to do any part of the analysis on your behalf.

Committing said acts can result in disciplinary action, which includes, but is not limited to failure on the assignment/course, disciplinary probation, or suspension. Substantial or repeated cases of misconduct will be forwarded to the Office of Student Conduct & Community Standards for additional review. For more information, refer to [students.wisc.edu/student-conduct/academic-integrity/](https://students.wisc.edu/student-conduct/academic-integrity/)