

CS Rubric – Age & Online Reviews

DS 4002 – Fall 2024 - Instructor: Loreto Alonzi – Kiera Murphy

Due: December 2024

**Submission format: Upload link to GitHub repo to Canvas and upload slide deck in PDF format
Individual Assignment**

General Description: Submit to canvas a link to your case study repository and a slide deck in pdf format

Preparatory Assignments – Read attached resources.

Why am I doing this? This assignment focuses on completing an analysis and creating a deliverable to discover how age may be an important factor in online reviews. You will be placing yourself in the shoes of a true data scientist and communicating your analytical findings in a way which is digestible for a non-data scientist.

- Course Learning Objective: use analysis and critical thinking skills.
- Course Learning Objective: effective presentation of findings.

What am I going to do? First, you should read the Hook Document to get an overview of the assignment. Then, read the resources provided on online review sentiment to gain a deeper understanding of the topic. Next, use the README and other resources available to you to obtain and clean data and perform analysis. Lastly, you will prepare a presentation on how you obtained your results and any significant findings that you had. Deliverables include:

- Github Repository – to provide all annotated code, figures, and resources so that your results may be replicated.
- Presentation – to guide the audience through your work.

Tips for success:

- Don't overthink it. Keep it simple. You are creating something for peers.
- Talk to your fellow students and instructors. It is important to receive feedback and ask questions of others.

How will I know I have Succeeded? You will meet expectations on this assignment when you follow the criteria in the rubric below.

Spec Category	Spec Details
Formatting	<ul style="list-style-type: none">● Repository – A GitHub repo containing all materials<ul style="list-style-type: none">○ Submit a link to the repo○ Everything is contained in the repo or linked to it if appropriate.

	<ul style="list-style-type: none"> ○ Contents <ul style="list-style-type: none"> ■ A README.md file (which auto displays) ■ A LICENSE.md file (use MIT as default) ■ A SCRIPTS folder ■ A DATA folder ■ AN OUTPUT folder ■ Presentation in pdf format ○ Use pdf format when possible ○ For code and data products use the appropriate format for whatever it is <ul style="list-style-type: none"> ● Presentation - should contain a 10-12 slide deck in pdf format
Github Repository	<ul style="list-style-type: none"> ● <u>Goal</u>: This repo is your credible method of explaining your results. Someone should be able to download this repo and produce the same findings and easily understand how to do so. ● README.md: explain how to use repo, sections designated using “##”. It should include a software and platform section, a map of your documentation, and instructions for reproducing your results. ● LICENSE.md: Select an appropriate license from the GitHub options list on repository creation. Usually, the MIT license is appropriate. ● SCRIPTS Folder: Include all the scripts you used. Try to name each script according to the order it needs to be executed to reproduce the results. All script files should include header comments at the beginning of a script to provide information that anyone working with or executing the script should be aware of. Throughout all your scripts, you should include copious comments explaining what each command or sequence of commands accomplishes and what the purpose is. ● DATA Folder: You should AT LEAST the data include the initial data, and the final data analyzed. If needed, the code in the SCRIPTS folder should be able to get you from the initial piece of data to the final one. N.B. If the initial and final data are the same, then just include that dataset. If your data fits in github, place all of it here. If your data does not fit in GitHub use a single file explaining the process to obtain the dataset. ● OUTPUT Folder: The content here can be in progress when MI3 is complete. It should be finished during MI4 though. Importantly, any information like tables, figures shown in your presentation should be here. Use informative names for your files. ● Copy of presentation in pdf format.
Presentation	<ul style="list-style-type: none"> ● <u>Goal</u>: This presentation is how you will primarily be communicating with your peers. It will allow you to communicate your findings clearly and efficiently. Include the following slides:

	<ul style="list-style-type: none"> ● Title & Outline (1 slide): Package your most important finding in headline format. Indicate the major parts of your presentation. ● Motivation/Context/Hypothesis/Research Question/Modeling Approach/Goal/Etc. (1-2 slides): Share relevant information with the audience so that they can understand what you are investigating. ○ Share relevant information so that they can understand why this matters. Restate Research Question. Restate Modeling Approach. ● Data Explanation/Acquisition (1 slide): Share the information from your data dictionary in a condensed way. Leave out most details but be prepared to answer follow-up questions. State the format (e.g.: text, image, etc.). State the size and necessary techniques if applicable. Discuss how your data set was acquired. Include relevant licensing or ethical concerns. ● Analysis Plan and Justification (1 slide): Show the diagram for your analysis pipeline. Highlight any particularly challenging components. Show any diagnostic or validation work done. ● Tricky Analysis Decision (1 slide): Every project has moments in analysis where you just need to make a decision, but the ‘right’ answer is not obvious or may not even exist. This slide is about one of those moments. Think about a judgment call you had to make in your project and describe why it matters and why it was tricky. Describe the impact of your decision. ● Bias and Uncertainty Validation (1 slide): List biases you consider and ones you correct for. Describe how you came to determine the uncertainty in the values you report. ● Results/Conclusions (1 slide): Address the hypothesis. Share relevant figures and findings ● Next Steps (1 slide): Describe new lines of exploration discovered during the project. Describes improvements that could be made. Describe New questions that have arisen. ● References/Resources/Acknowledgements (1 slide): Provide citations as necessary. Provide links to resources used (e.g.: your GitHub page). Acknowledge contributions and assistance ● Closing Slide (1 slide): This slide will sit up while people are asking questions, pick something memorable. It can be a repeat from earlier. Don’t be afraid to return to an earlier slide if a question takes you there.
--	---

Acknowledgements: Special thanks to Jess Taggart from UVA CTE for coaching on making this rubric.
This structure is pulled from [Streifer & Palmer \(2020\)](#).