Yelp Sentiment Analysis Case Study Rubric

DS 4002 – Spring 2025 – Adaire Burnsed
Due: May 9, 2025
Submission format: GitHub repository (submitted by link to canvas)

Individual Assignment

General Description: Submit to Canvas a link to your GitHub Repository for this case study

Why am I doing this? This assignment is your opportunity to carry out the analysis plan outlined in the case study that you selected and generate results that directly support (or refute) your case study's core question. You're not just building toward a final presentation, you're practicing the full research process, from working with real data to documenting your methods clearly and professionally. It's easy to rush through the steps and focus only on results, but this assignment emphasizes both the process and the product. By carefully documenting your work and following through on each stage of analysis, you'll gain experience that mirrors real-world data workflows.

What am I going to do? In this individual case study, you'll take full ownership of the analysis from start to finish. You'll apply the research plan you've developed by following a structured workflow and documenting your process clearly. This assignment is not only about producing results, it's about demonstrating a thoughtful and professional approach to data analysis. You will begin by following the plan laid out by the case study and work your way towards creating a presentation with your results. Throughout this process, you'll need to make careful decisions about your methods and ensure your documentation is clean, reproducible, and clearly communicates your work. Your final goal is to create a case study presentation that tells a compelling, data-driven story about the connection between restaurant price and review sentiment.

Tips for success:

- Focus, focus, focus – Avoid distractions (e.g. put the cellphone down and turn off notifications).
- Focus, focus, focus – Research shows that if you are programming and you have a distraction it takes you 20 minutes to get your head back into the state it was in prior to the start of the distraction. Find a way to block out your time and your mind and you can get quite a lot done in a short focused window.
- Temper curiosity – this assignment is not the time to explore every last nook and cranny. If something interesting pops up, write it down but don't pursue it if it is not on the path to your goal.

How will I know I have succeeded? You will meet expectations on CS3 when you follow the criteria in the rubric below.

[redacted]

Formatting	• One Github Repository (submitted via link on canvas)
	• To ensure **reproducibility**, the repository will adapt parts of the TIER Protocol 4.0. In a nutshell, the top level page of the repository should contain:
	  ○ A README.md file (which auto displays)
	  ○ A LICENSE.md file (use MIT as default)
	  ○ A SCRIPTS folder
	  ○ A DATA folder
	  ○ AN OUTPUT folder

README.md	• Goal: This file serves as an orientation to everyone who comes to your repository, it should enable them to get their bearings.
	• Use markdown headers to divide content.
	• Make an H2 (##) section explaining the contents of the repository
	• Section 1: Software and platform section
	  ○ The type(s) of software you used for the project.
	  ○ The names of any add-on packages that need to be installed with the software.
	  ○ The platform (e.g., Windows, Mac, or Linux) you used.
	• Section 2: A Map of your documentation.
	In this section, you should provide an outline or tree illustrating the hierarchy of folders and subfolders contained in your Project Folder, and listing the files stored in each folder or subfolder.
	• Section 3: Instructions for reproducing your results.
	In this section, you should give explicit step-by-step instructions to reproduce the Results of your study. These instructions should be written in straightforward plain English, but they must be concise, but detailed and precise enough, to make it possible for an interested user to reproduce your results without much difficulty. **N.B. This section will be crucial for the CS1 assignment, where you'll be required to reproduce the results of other groups. Therefore, make sure to explain this section thoroughly.**

LICENSE.md	• Goal: This file explains to a visitor the terms under which they may use and cite your repository. • Select an appropriate license from the GitHub options list on repository creation. • Usually, the MIT license is appropriate.
SCRIPTS folder	• Goal: This folder contains all the source code for your project. • 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	● Goal: This folder contains all of the data for this project. ● 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. ● A Data Appendix file as a PDF, which will include text that you type, as well as tables, figures, and other descriptive statistics. This file should be organized in sections, with a section for **each** dataset analyzed. Each section should begin with a statement of what the unit of observation is--that is, it should explain what kind of object each row of the data file represents. After that, you should include a subsection for each variable in the analyzed dataset. More information: https://www.projecttier.org/tier-protocol/protocol-4-0/root/data/analysisdata/data-appendixfile
OUTPUT folder	● Goal: This folder contains all of the output generated by your project, e.g. figures, tables, etc. ● 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**.
Written Reflection	● Goal: This file should summarize your key findings and takeaways from the case study analysis in a clear and concise way. ● One page (pdf) ● Things to discuss include: how your analysis worked, your results, challenges and/or big successes, and a reflection on your work in general.
References	● All references should be listed at the end of the document ● Use IEEE Documentation style (link)

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