Yelp Case Study Rubric

DS 4002 – Spring 2024 - Creator: Ananya Goel
Due: May 11, 2025
Submission format:

- **Upload link to a github repository with all relevant material to replicate this project**
- **Upload presentation in PDF format**

Individual Assignment

Preparatory documents: Study prompt, articles

Why am I doing this? The goal of this assignment is to successfully conduct a sentiment analysis on Yelp reviews using the dataset provided. You will learn how to work with JSON tables, the VADER package in python, and the ANOVA test. But remember, presenting your results in a clear and concise manner is more important than conducting the analysis. So consolidate your results and be ready to explain your work as a presentation.

- Course Learning Objective: Prepare findings for presentation to your peers.

What am I going to do? You are going to use the data and guidelines provided in the GitHub link to perform a sentiment analysis. Use the articles provided to guide your project and motivation for this project. You will need to combine JSON tables and convert them into a usable csv. You will want to clean your data depending on what your scope is. You should familiarize yourself with the VADER package as you will use it to conduct the sentiment analysis.

- GitHub repository link with relevant code and output
- Presentation with results and conclusions in PDF format

Tips for success:

- Do not be afraid to make mistakes. Use all resources provided to you and feel free to debug any errors you run into online or by contacting the creator of this case study.

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

Formatting	• One Github Repository (submitted via link on canvas)
	  ○ To ensure **reproducibility**, the repository will adapt parts of the TIER Protocol 4.0.
	  ▪

	● Presentation- Explaining the project, motivation, and results o PDF format
GitHub Repo	● A README.md file (which auto displays) ○ Topic, explanation/motivation, what is included in the Repo, how to reproduce project ● A SCRIPTS folder o This folder contains all the source code for your project. o Name each script according to the order it needs to be executed to reproduce the results and comment it well ● A DATA folder o This folder contains all of the data for this project. o It should include the initial data, and the final data analyzed ● AN OUTPUT folder o This folder contains all of the output generated by your project, e.g. figures, tables, etc. ○ Include your final presentation here
Presentation	● About 8-10 slides ● PDF format ● Slide numbers (except for title slide) ● Order o Title & Outline o Motivation/Context/Hypothesis/Research Question/Modeling Approach/Goal/Etc. o Data Explanation/Acquisition o Analysis Plan and Justification o Tricky Analysis Decision o Bias and Uncertainty Validation o Results/Conclusions o Next Steps o References/Resources/Acknowledgements o Closing Slide
References	● All references should be listed at the end of the document ● Use IEEE Documentation style (link)

Acknowledgements: Special thanks to DS 4002 professors for coaching on making this rubric. This structure is pulled from Streifer & Palmer (2020).