CASE STUDY 3 RUBRIC — SENTIMENT ANALYSIS OF FILM CRITIQUES

DS 4002 – Fall 2024, Mina Gorani
Due: December 9
Submission Format: GitHub Repository
General Description: Submit to canvas a link to your GitHub repository.

Why am I doing this? Often, questions surface about social trends and norms on the Internet. This is your opportunity to harness your technical and analytical skills to apply textual sentiment analysis and help answer one of these questions: Do movie critics have a backbone? In this scenario, you will not only need to apply sentiment analysis, but also develop the quantitative benchmarks that you deem are sufficient markers of the social trend.

What am I going to do? You have been provided with a Project Source Repository with two datasets: one with movie information and another with movie critiques pulled from Rotten Tomatoes. With these two datasets, you will clean and familiarize yourself with the datasets. Then, analyze if there is a shift in movie critic's critique sentiments before and after a film is released for streaming. Then, you will further analyze if this shift increases or decreases alignment with the sentiments of audiences. Lastly, you will produce visuals to communicate the results of your analysis. A template python script is also provided in the Project Source Repository which guides you through the basic steps of this project.

Tips for success:

- Follow your curiosity — the resources for this project are intentionally open ended. You'll get the most out of this experience if you explore the questions that naturally surface as you read through the sources and familiarize yourself with the dataset.
- Don't be scared of trying something new — the templates in the provided scripts are just guidance. If you have another idea for how to approach analysis or visualization, try it out! Your instructors and generative AI are fantastic resources to help you explore new methods.

Spec Category	Spec Details

Formatting	• Repository – A GitHub repo containing all modified or produced materials
	  ○ Submit a link to repo
	  ○ README.md: a general explanation of the project and materials included in the repo
	  ○ LICENSE.md file (use MIT as default)
	  ○ Folders
	  ▪ Data: raw data as provided in the project source repo
	  ▪ Scripts: your modified python notebook (.ipynb) from the project source repo
	  ▪ Output: All graphs/visuals (.png or .jpg) and slidedeck communicating findings (pdf)

README.md	• Goal: Orient visitors to your repository
	• Use markdown headers to divide content including
	  ○ (1) Software & platform section with the software and any add-on packages you used (including those already added to the template python notebook)
	  ○ (2) Map of documentation with an outline of the files and folders in your repo
	  ○ (3) Instructions for reproducing your results, highlighting your additions to the template python notebook
	  ○ (4) References

LICENSE.md	• Goal: This file explains to a visitor the terms under which they may use and cite your repository • Select the MIT license from the GitHub options list on repository creation

Modified Python Notebook (SCRIPTS folder)	• Goal: This folder will contain all scripts used to get the raw data to the form of the outputs
	• Using the provided python notebook (found on the GitHub Repo linked on the hook document) as a template, clean the source data, carry out sentiment analysis, and produce visuals to demonstrate results
	  ○ Upload your modified notebook to your repo
	  ○ Don't limit your contributions just to the cells with commented directions. Feel free to add in cells, reorder processes, or further analysis

DATA folder	• Goal: This folder contains all the data used in this project • Place the directions to obtain the raw data in the Project Source Repository into this folder along with any other data you may have used • Include a Data Appendix file as a PDF organized by each dataset analyzed ○ Include tables, figures, and descriptive statistics

Slidedeck (OUTPUT folder)	• Goal: This folder contains the output generated by your project that should communicate your findings
	  ○ Include all tables and figures produced by your scripts
	  ○ Create a slidedeck walking an audience through your findings. At minimum this should include
	  ▪ Introduction/Hook sharing the nature of this exploration
	  ▪ Research Question defining the quantitative benchmark you set to explore the subject
	  ▪ Limitations including biases or assumptions
	  ▪ Results including output visuals from your script
	  ▪ Next steps
	  ▪ References

References	• All references should be listed at the end of your README.md and slidedeck in IEEE format