Movie Critic Case Study Rubric

DS 4002 – Fall 2024 – Anagha Chundury
Due: Dec 9
Submission format: Upload link to GitHub repository on UVA Canvas

Why am I doing this?
This is your opportunity to synthesize your data science knowledge thus far and showcase your technical and analytical skills. This case study represents an example of a real life data science quandary that you may be tasked with exploring in the future.

What am I going to do?
You will establish an analysis plan and carry it out. You will acquire the necessary datasets (available on the github repository) which include a dataset of critics reviews and associated movies, a dataset of movies and their associated release to streaming platforms, and a dataset containing all of the audience rotten tomatoes scores. You will combine the datasets and use the vADER sentiment package in Python to conduct the sentiment analysis. The analysis will involve interpreting sentiment scores within the range of -1 (most negative) to 1 (most positive), with scores falling between -0.05 and 0.05 considered as indicative of neutral sentiments. The sentiment scores before and after the streaming release date will be compared in addition to the RT already available RT Audience score.

Your final deliverables should include:
- A document detailing your analysis plan and results
- Source code
- A github repository with all of your sources and materials

Tips for success:
- Keep track of all of your documentation so you can reference it easily!
- When thinking of how to compare scores, don't be afraid to scale up or scale down so they are comparable. You have creative liberty!

How will I know if I have succeeded?
You will meet expectations on this case study when you successfully follow and complete the criteria in the rubric below:

Spec Category	Spec Details
Formatting	- One GitHub repository submitted via link on Canvas - The top level page should contain:

	○ README.md
	   ■ A brief summary of what you have produced for the case study
	○ LICENSE.md
	   ■ Select an appropriate license so visitors know how to interact with your repository
	○ Source Code File
	○ Analysis File
	○ References File

Document of Analysis Plan + Results	Discuss your results and analysis by addressing these points: ● Summarize the problem you were tasked with solving ● Describe the steps of your analysis plan ● Provide the visualizations of your results you have created ● Include a discussion which explains your results and references your visualization ● Discuss challenges and future steps
Source Code	Your source code should include: ● An exploratory data analysis which explores questions such as the number of movies and the number of reviews per movie ● A sentiment analysis of the movie critics reviews before and after ○ This should be done using the data files labeled "rotten_tomatoes_movie_reviews" and "rotten_tomatoes_movies" ● Comments which clearly detail what that section is for in reference to your analysis plan
References	● Include a list of references used for the project *in addition to* the already provided references