CS3 - Sentiment Score Analysis Case Study Rubric

DS 4002 - Fall 2024 - Taylor Toth
Submission Format: Presentation and Github Repository

Purpose:
Completing this project will allow for investigation and further practice your text data analysis techniques using a real world application. The goal of this assignment is to create a sentiment analysis model that investigates movie review and Bechdel-Test data from the recent decades that gives insight into movie trends based on evolving gender representation. This will be done using a VADER model to create sentiment scores from IMBd movie reviews from the recorded most popular movies. By the end of this study, you will have cleaned the data, performed an early exploratory analysis, and created an analytical model using real life data.

Task:
Navigate the materials provided in the Github repository in order to follow the process of retrieving the sentiment scores and analyzing the data using the provided Python scripts. The final version of this project will be a self-typed report and presentation that could be presented to the company directors about gender based trends over time and reasoning on why this information is important to the future movie selection process. Your report should be supported by visualizations and other figures that help aid in the understanding of the analysis to help facilitate future decisions.

Tips for Success:

- Avoid lengthy explanations that could make the information difficult to digest.
- Do background research in cultural changes over time that could have an impact on gender representations in film to know where possible significant results could be found.
- Create lots of early and conclusion based visualizations to support findings for easier understanding of the data and results.
- Be clear, direct, and intentional in your conclusion.

How will I know I succeeded?

At the end of the project, you should have a professional presentation that summarizes all of the relevant trends and gives recommendations for the movie types that would perform best on Netflix in the future. This is to include visualizations, example code, and meaningful conclusions. Onward!

Spec Category	**Spec Details**
Repository and Submission	Goal: You will submit a link to your GitHub Repository that contains the following elements: ● All data, including precleaned and all modified versions ● Early exploratory visualizations ● All scripts which should be well commented with headers ● Folders sorting the data based on visualizations, scripts, and any other important materials used ● Creation of a READme and listening file ● All references must be included
Sentiment and Bechdel Test Score Preparation	Goal: Successful execution of scripts needed to scrape websites for movie data regarding the Bechdel test and IMBd reviews for sentiment score creation: ● Execute ScrapeNewBechdel.py and IMDbReviewSentiment.py Clearly document the steps used to create the VADER model in the IMDb sentiment document to provide a clear understanding of the produced sentiment scores.
Exploratory Data Analysis	Goal: Successful execution of scripts needed to generate visualizations based on the Bechdel Test and sentiment score data that was just scraped: ● Execute ExploratoryPlots.py Create professional and clean-looking plots that reveal early trends in the data. Discuss these trends and any possible routes of exploration.
Data Preparation and Cleaning	Goal: Successful execution of scripts needed to transform the data and perform the statistical analysis:

	● Execute RoundSentiment.py and AddBinaryRating.py. Explain the changes made to the data and why these changes were integral to creating the cleaned data set.
Modeling & Evaluation	Goal: Successful execution of scripts needed to perform the logistical regression, hypothesis testing models, and conclusions supported by these models: ● Execute LogisticRegression.py, HypothesisTesting.py, and AnalysisPlots.py ● Optional execution of HypothesisTestingTime.py Successfully explain why these models were chosen and how they correlate to your projects motivation. Provide clean visualizations and concise yet meaningful results that explain results. Discuss the model's overall performance and why you are confident in your model's power.