Case Study Rubric – Chick Flicks and Data Tricks

DS 4002 – Spring 2024 - Instructors: Mary Ellen Schuster
Submission format: Upload link to github repo to canvas

Individual Assignment

General Description: Submit to canvas a link to your case study repository

Preparatory Assignments – CS1 and CS2 should help you better understand how this case study should be completed.

Why am I doing this? You are undertaking this project to explore the intersection of data science and real-world applications within the media and entertainment industry. This case study is designed to enhance your analytical skills, familiarize you with advanced sentiment analysis tools, and demonstrate how data-driven insights can be used to predict and influence public perceptions and outcomes.

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

What am I going to do? To begin, read the rubric and the deliverable in their entirety. Research types of sentiment analysis tools available to you for the project. Gather and preprocess a dataset of IMDB reviews for the Barbie movie, and use your chosen tool to analyze the reviews and extract sentiment scores. Create a model to correlate sentiment scores with IMDB ratings, training, testing, and evaluating it for accuracy. Finally, compile your results and overall insights into a presentation.

Tips for success:

- Do your Research: Be sure you are very familiar with the sentiment analysis model you choose. Research how this model has been used in the past, and whether it is applicable to the type of data you are using.
- Ask for help: Be sure to reach out to the professors if you need it!
- Have fun: Take liberties with this project based on where the data takes you. Be creative, and enjoy the data analysis process!

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

Formatting	● Repository – a Gitbhub repository containing all materials

	○ 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 ● Presentation ○ About 7 slides ○ PDF format for submission to collab ○ Generate the slides through the program of your choice ○ Slide numbers (except for title slide)
Github Repository	● Goal: Explain the purpose of your model and the overall findings ● Contents: ○ README.md: discusses project overview, background research, links, etc ▪ Section 1: Software and platform section: The type(s) of software you used for the project. ▪ Section 2: A Map of your documentation ▪ Section 3: Instructions for reproducing your results. ○ DATA: this folder includes original source data ○ SCRIPTS: This folder contains all the source code for your project. ○ OUTPUT: This folder contains all of the output generated by your project, e.g. figures, tables, etc. ○ LICENSE: This file explains to a visitor the terms under which they may use and cite your repository.
Presentation	● Goal: Display the analysis and findings in an easy-to-digest manner. Should be entertaining and eye-catching. ● Presentation ○ Order ○ Title & Outline ○ Motivation/Context/Hypothesis/Research Question/Modeling Approach/Goal/Etc.

	○ Data Explanation/Acquisition ○ Analysis Plan and Justification ○ Tricky Analysis Decision ○ Bias and Uncertainty Validation ○ Results/Conclusions ○ Next Steps ○ References/Resources/Acknowledgements ○ Closing Slide
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).