Case Study - Comparing Public Sentiment Between Barbie vs. Oppenheimer

DS 4002 – Spring 2024 - Colby Le

Submission format:

- **PDF document, source code**

Individual Assignment

General Description: PDF document, source code

Prerequisites: exposure to the Barbenheimer trend, as well as general programming skills

Why am I doing this? The goal of this assignment is that it proves sentiment analysis can be leveraged to answer seemingly subjective questions using real objective data. It proves movie comparisons as well as general public popularity contests can be settled by purely qualitative data that can be converted into quantitative, observable values. In addition to this, this assignment will prove that you are capable of implementing a real coding solution for an abstract question, and that you are able to learn the intricacies of your model and then report that to a wider audience.

- Course Learning Objective: understand the application of natural language processing in subjective questions.
- Course Learning Objective: prepare findings for presentation to your peers.

What am I going to do? Produce a coding solution and report for the Warner Bros Movie Studio as a contracted data scientist for the purpose of grasping public opinion about the ultra-profitable Barbenheimer phenomenon of Summer 2023. This will include a refined Github repository containing your data and solution notebook, and in addition a brief typed report containing an executive summary of your solutions, a description of the project's motivations to serve as persuasion to studio executives, as well as your conclusions. You are encouraged to use graphs and figures to communicate your findings.

- Github repository containing your data and dsolution code, as well as any data analysis and data visualizations you wish to produce
- 2-3 page summary document for studio executives

Constraints:

- If you feel like the data is not satisfactory in any way, you can drop entries and clean the data as you wish.
- You are required to reach a conclusion about which movie was more popular (Barbie or Oppenheimer)

Tips for success:

- Study the topic of public sentiment before you attempt the project.
- Be careful about the public model you select to perform the sentiment analysis.

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

Board Room Report	● Goal: produce a report for the studio executives that includes an executive summary of your solutions, a description of the project's motivations to serve as persuasion to studio executives, as well as your conclusions. ● Formatting: 2-3 pages, PDF Format ● Executive Summary: o Introduce the solution. o Describe the process you underwent to get to that conclusion. Use as much non-technical language as possible but also acknowledge the technology you used. ● Motivation and Context: o Describe the exigent need for the answer to the question "which movie was better" - describe the millions of dollars at stake and how using an AI solution to problems like these can be very profitable to the movie studio. ● Conclusions: o Present the solution clearly: which movie was better, and on what basis? o Describe the quantifiable data you collected that supports your conclusions.
Github Repository	● Goal: produce an easily navigable repository that contains the code for your project ● It should contain: o README.md: describe the contents of the repository, including the file tree, context, and contact information for each of the project o SCRIPTS folder: contains the notebooks which you perform data cleaning, exploratory data analysis, and the actual sentiment analysis in. o OUTPUT folder: contains any data visualization and results your notebooks generate. o DATA folder: contains data you sourced.