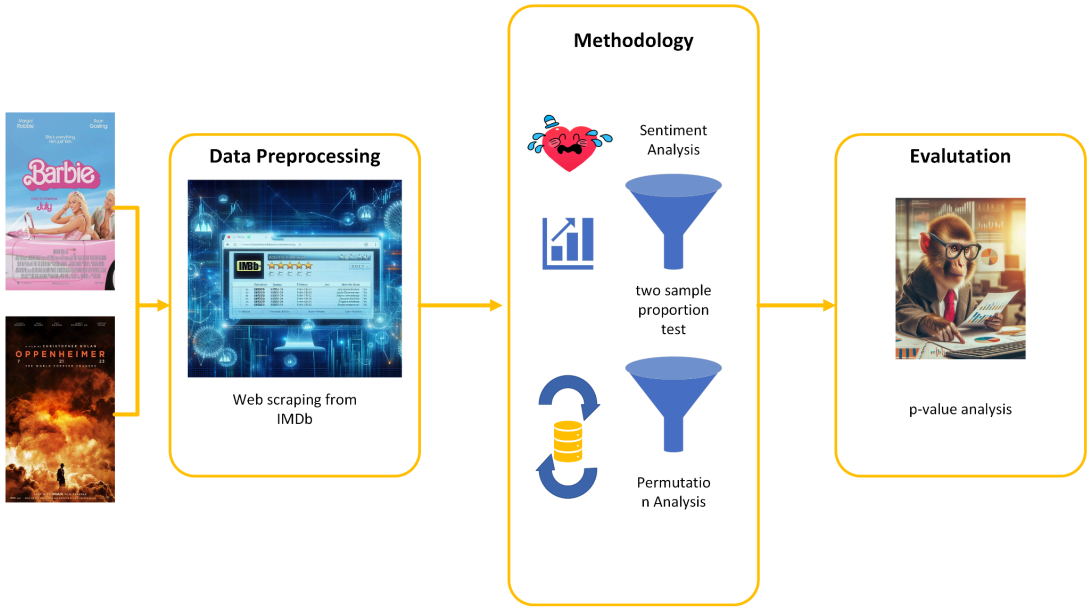
Professor Javier Rasero

May 10, 2024

Comparing Public Sentiment Between Barbie vs. Oppenheimer

In the era of data-driven decision-making, understanding public sentiment towards movies is crucial for movie studios aiming to maximize their profit margins. Sentiment analysis, a subfield of natural language processing, offers a powerful tool to quantify the positive or negative sentiments expressed in movie reviews. For example, a review such as “I did not enjoy this movie one bit” would be classified as strongly negative sentiment by a model. But high-performing models must be able to make the distinction for more subtle statements, such as “This movie ran a little long for my taste,” which does not contain outwardly negative words but would be classified as with a moderately negative sentiment score. It is up to you to select more nuanced pre-trained models which are not limited by high word counts which can be found in some reviews, and can make more precise sentiment estimations beyond simply “positive” or “negative.” Pre-trained models can be found at https://huggingface.co/docs/hub/en/transformers.

In this report, we analyze the public sentiment towards the movies Barbie and Oppenheimer using sentiment analysis techniques. Our goal is to provide the studio executives with actionable insights to inform their decision-making process for the upcoming summer movie season.



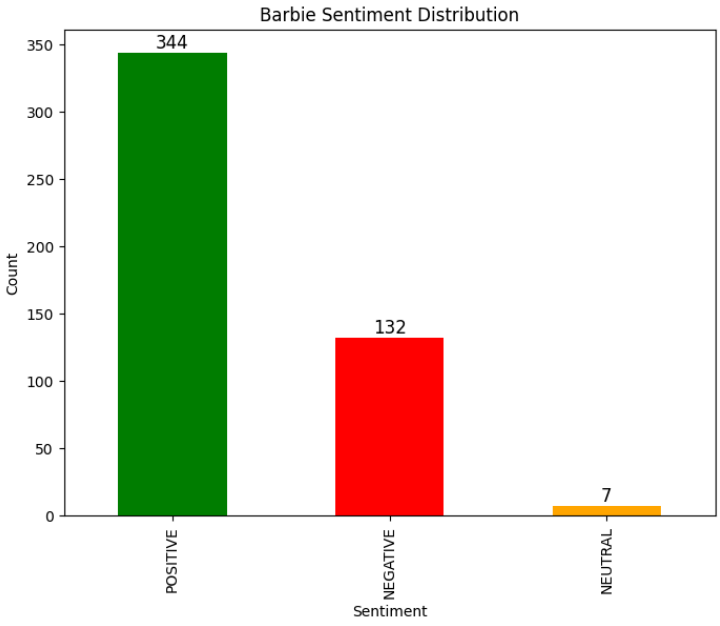
Methodology:

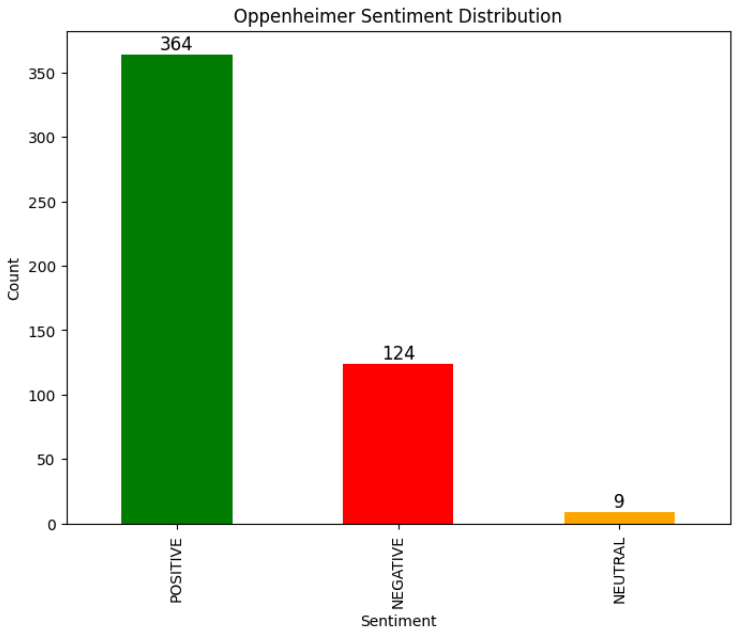
The following is an example methodology that can be performed, based off of our own methodology for the project. IMDb provides a rich source of user-generated reviews, allowing us to gauge public sentiment effectively. Our approach involved the following steps:

- Data Collection
    - We retrieved movie reviews from IMDb for both Barbie (https://www.imdb.com/title/tt1517268/reviews) and Oppenheimer (https://www.imdb.com/title/tt15398776/reviews).
    - This requires some web-scraping solution, for example Selenium
- Preprocessing
    - Clean the data by dropping reviews which are longer than our selected model's word limits. We also performed tokenization to break down the reviews into individual words or phrases.
    - You are also free to target only specific time frames for the reviews, as IMDb provides that information. In our case, we only looked at Summer 2023 reviews.
- Sentiment Analysis
    - We employed a pre-trained sentiment analysis model to classify each review as positive, negative, or neutral based on the overall sentiment expressed.
- Aggregation
    - We aggregated the sentiment scores of individual reviews to calculate an overall sentiment score for each movie. This can be as simple as summing every sentiment score for each review per movie, then dividing by the number of reviews which belonged to that movie, though more statistical measures exist.

Results:

In our own results, after analyzing a significant number of reviews for both Barbie and Oppenheimer, we found the following sentiments:





Our definitions of a “neutral” score are loosely subjective. If the model indicated a confidence score for its decision less than 0.7, we considered that a neutral sentiment.

Average Sentiment Score

Barbie	Oppenheimer
0.4244306418219462	0.4647887323943662

Based purely on the average sentiment score, Oppenheimer was the more liked movie. Provide more justification and metrics for your own takeaway!

Recommendations:

An example opinion we could recommend to the Warner Bros executives was that Oppenheimer was the better performing movie. From there, they could look at the factors behind that movie, or closely analyze the reviews to see what the audience preferred in that movie over Barbie. The main takeaway should be that studios should not blindly look at the numbers, such as ticket sales, to understand what their audiences like the most.