# barbie_reviews.csv

This dataset contains web-scraped IMDb reviews for “Barbie” (2023). Each row represents an individual review from the website.

| **FEATURE NAME** | **DATA TYPE** | **DESCRIPTION** |
|---|---|---|
| score | String | Rating out of 10 the reviewer gave this movie. Can be compared against the final sentiment scores. |
| title | String | The title the reviewer labeled their review. |
| date | Int | Month-year of the review in the form XX XXXX. Allows us to filter by date for the first three months of reviews. |
| text | String | Long-form string, actual contents of the review. |

# oppenheimer_reviews.csv

This dataset contains web-scraped IMDb reviews for “Oppenheimer” (2023). Each row represents an individual review from the website.

| **FEATURE NAME** | **DATA TYPE** | **DESCRIPTION** |
|---|---|---|
| score | String | Rating out of 10 the reviewer gave this movie. Can be compared against the final sentiment scores. |
| title | String | The title the reviewer labeled their review. |
| date | Int | Month-year of the review in the form XX XXXX. Allows us to filter by date for the first three months of reviews. |
| text | String | Long-form string, actual contents of the review. |

# barbie_reviews_sentiment.csv

This dataset is an extension of the web-scraped Barbie review data, with sentiment and sentiment_score features added post-sentiment analysis.

| **FEATURE NAME** | **DATA TYPE** | **DESCRIPTION** |
|---|---|---|
| score | String | Rating out of 10 the reviewer gave this movie. Can be compared against the final sentiment scores. |
| title | String | The title the reviewer labeled their review. |
| date | Int | Month-year of the review in the form XX XXXX. Allows us to filter by date for the first three months of reviews. |
| text | String | Long-form string, actual contents of the review. |
| sentiment | String | NEGATIVE, POSITIVE, |

| | | |
|---|---|---|
| | | NEUTRAL. The sentiment analysis process assigns NEGATIVE and POSITIVE to each review, and if the sentiment/confidence score is less than 0.7 then assign NEUTRAL. |
| sentiment_score | Int | Numerical encoding of the sentiment: 1 if POSITIVE, 0 if NEUTRAL, -1 if NEGATIVE. |
| cluster | Int | 0, 1, 2 corresponding to sentiment_scores 1 (POSITIVE), 0 (NEUTRAL), -1 (NEGATIVE). This designates the assigned cluster for the review in the sentiment clustering. |

# oppenheimer_reviews_sentiment.csv

This dataset is an extension of the web-scraped Oppenheimer review data, with sentiment and sentiment_score features added post-sentiment analysis.

| **FEATURE NAME** | **DATA TYPE** | **DESCRIPTION** |
|---|---|---|
| score | String | Rating out of 10 the reviewer gave this movie. Can be compared against the final sentiment scores. |
| title | String | The title the reviewer labeled their review. |
| date | Int | Month-year of the review in the form XX XXXX. Allows us to filter by date for the first three months of reviews. |
| text | String | Long-form string, actual contents of the review. |
| sentiment | String | NEGATIVE, POSITIVE, |

| | | |
|---|---|---|
| | | NEUTRAL. The sentiment analysis process assigns NEGATIVE and POSITIVE to each review, and if the sentiment/confidence score is less than 0.7 then assign NEUTRAL. |
| sentiment_score | Int | Numerical encoding of the sentiment: 1 if POSITIVE, 0 if NEUTRAL, -1 if NEGATIVE |
| cluster | Int | 0, 1, 2 corresponding to sentiment_scores 1 (POSITIVE), 0 (NEUTRAL), -1 (NEGATIVE). This designates the assigned cluster for the review in the sentiment clustering. |