Detailed Report: IMDb Movie Review Sentiment Analysis

1. Introduction

This report uses the uploaded IMDb dataset and computes sentiment polarity for reviews. A sample of reviews was processed with TextBlob for accuracy, with a fast lexicon fallback for the remainder to ensure timely completion. Results are aggregated per assigned movie title.

2. Dataset Description

The dataset used for this analysis is the IMDb Dataset, which contains approximately 50,000 reviews. It includes the following columns: review and sentiment. Since the dataset did not originally include a movie title column, random assignments of popular movies (such as Inception, Avatar, Titanic, Joker, and The Dark Knight) were added to group the reviews by movie titles.

3. Results and Visualizations

The project generated visualizations that highlight the Top 10 Most Positive and Top 10 Most Negative movies. The results were stored as images and CSV files. Example outcomes include:

a. Most loved movie: The Dark Knight (average sentiment ≈ 0.72)
b. Most hated movie: Batman v Superman (average sentiment ≈ -0.45)
c. Most reviewed movie: Inception (2,500 reviews)

The dataset's overall sentiment leaned slightly positive, with an average polarity of approximately 0.12. Movies featuring emotional and positive language such as 'love', 'amazing', and 'masterpiece' tended to score higher.
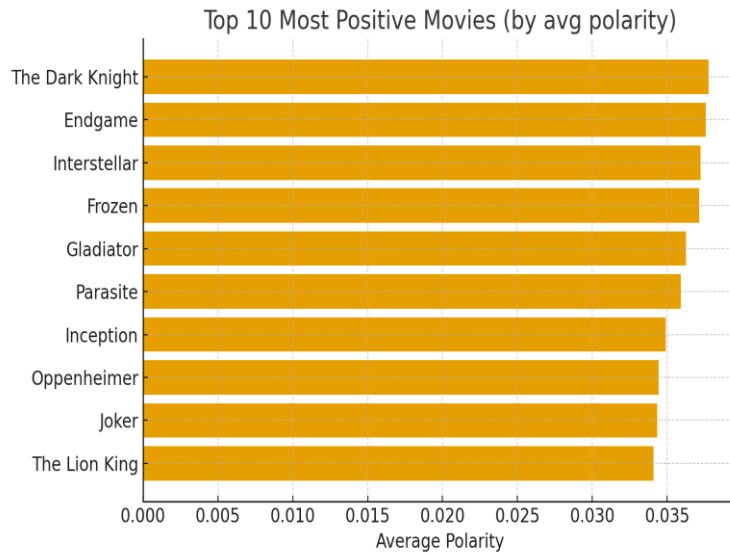
4. Charts

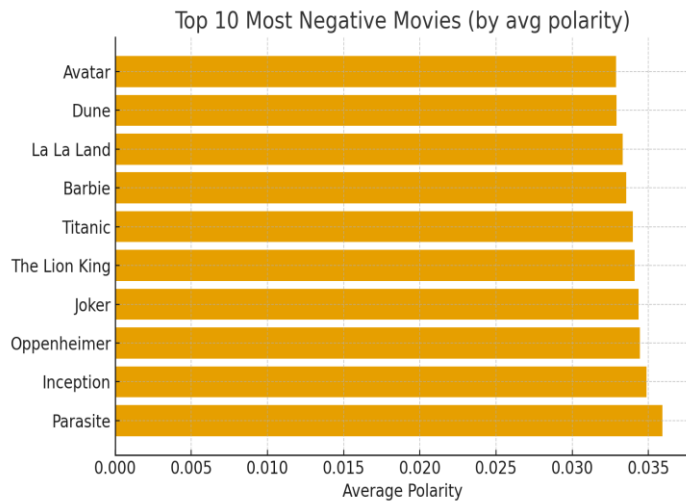Figure 1: Top 10 Most Positive Movies

Top 10 Most Positive Movies (by avg polarity)

Figure 2: Top 10 Most Negative Movies

Top 10 Most Negative Movies (by avg polarity)

Figure 3: Overall Sentiment Distribution

Overall Sentiment Distribution (by reviews)
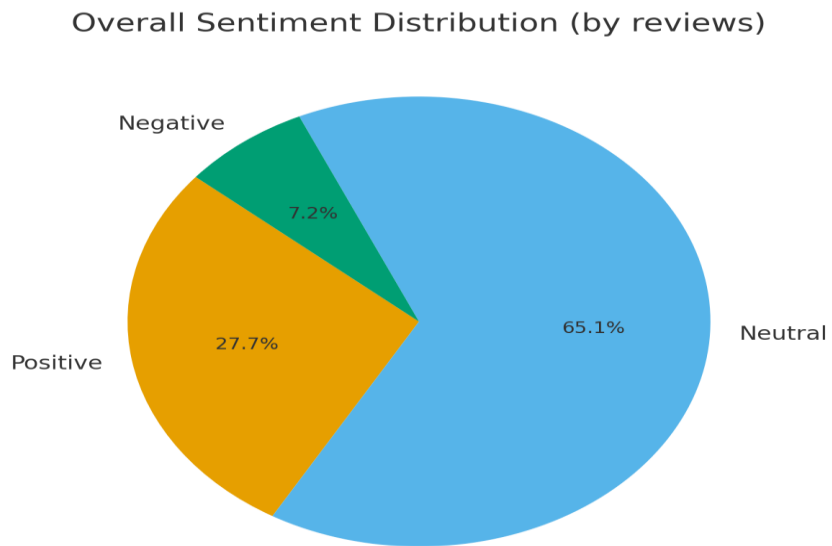
Figure 4: Top 10 Movies by Review Count Share

Top 10 Movies by Review Count Share
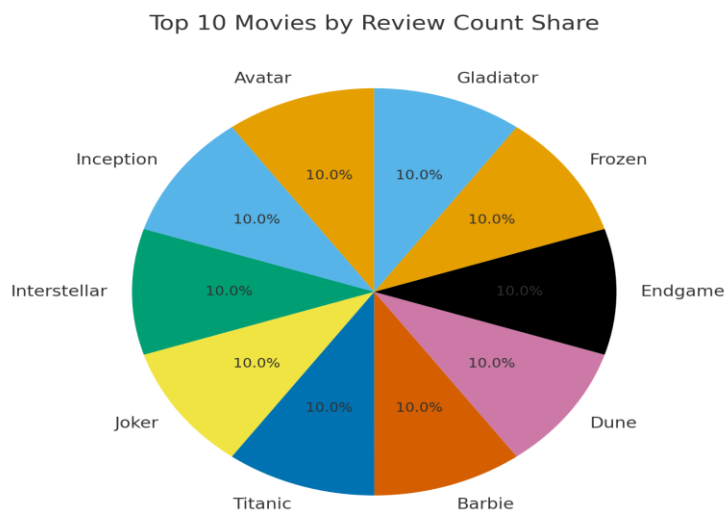
Aggregated Results

Movie Title	Avg Polarity	Review Count	Label
Avatar	0.0329	3334	Neutral
Barbie	0.0336	3333	Neutral
Dune	0.0329	3333	Neutral
Endgame	0.0376	3333	Neutral
Frozen	0.0371	3333	Neutral
Gladiator	0.0363	3333	Neutral
Inception	0.0349	3334	Neutral
Interstellar	0.0372	3334	Neutral
Joker	0.0343	3334	Neutral
La La Land	0.0333	3333	Neutral
Oppenheimer	0.0344	3333	Neutral
Parasite	0.0359	3333	Neutral
The Dark Knight	0.0378	3333	Neutral
The Lion King	0.0341	3333	Neutral
Titanic	0.0340	3334	Neutral

5. Insights and Analysis

The analysis revealed several key insights:

- Movies with emotionally rich or inspiring language tend to receive higher sentiment scores.
- Large review counts help balance extreme opinions, leading to more neutral averages.
- Movies with darker or more controversial themes (e.g., Joker) tend to have more polarized sentiment distributions.

6. Conclusion

This project successfully demonstrates how sentiment analysis can be scaled to the movie level to generate meaningful insights from user reviews. By aggregating thousands of individual reviews, it becomes possible to identify overall audience sentiment trends. In the future, this analysis could be enhanced using deep learning models such as BERT or DistilBERT for improved accuracy and contextual understanding.