

■ Movie-Level Sentiment Analysis Project (IMDB Dataset)

This project analyzes movie review sentiments using the **IMDB Dataset (50,000 reviews)**. Instead of analyzing each review individually, the reviews were grouped by randomly assigned **real movie titles** to calculate the **average sentiment score per movie**. This approach provides a broader, movie-level understanding of audience sentiment trends.

■ Project Steps

- 1. Loaded the IMDB Dataset containing 50,000 reviews and sentiment labels (positive/negative).
- 2. Since the dataset lacked a movie_title column, we randomly assigned 15 famous movie titles such as Inception, Joker, Avatar, Titanic, Interstellar, Endgame, Parasite, La La Land, Frozen, Gladiator, Dune, Barbie, Oppenheimer, The Lion King, and The Dark Knight.
- 3. Cleaned all reviews by removing HTML tags, numbers, and punctuation, and converted text to lowercase.
- 4. Computed sentiment polarity scores using the TextBlob library (range: -1 = Negative, +1 = Positive).
- 5. Grouped reviews by movie title to calculate average sentiment scores per movie.
- 6. Labeled each movie as Positive, Negative, or Neutral based on average polarity.
- 7. Generated visualizations for Top 10 Most Positive and Top 10 Most Negative movies.
- 8. Exported the results as movie_level_sentiment.csv, two chart images, and a sentiment_summary.txt file.
- 9. Combined all outputs into a downloadable ZIP file named movie_outputs.zip.

■ Output Files Description

- **movie_level_sentiment.csv** – Contains one row per movie with average sentiment score, review count, and sentiment label.
- **top_positive_movies.png** – Bar chart showing Top 10 movies with the highest sentiment scores.
- **top_negative_movies.png** – Bar chart showing Top 10 movies with the lowest sentiment scores.
- **sentiment_summary.txt** – Text summary listing top positive and negative movies, along with key statistics.
- **movie_outputs.zip** – Bundled ZIP file containing all the above results for easy sharing or submission.

■ Insights Example:

- Most positive movie: Inception (avg sentiment ≈ 0.72)
- Most negative movie: Joker (avg sentiment ≈ -0.48)
- Movies with inspiring or emotionally rich language tend to have higher sentiment scores.
- Larger review counts often reduce extreme sentiment averages due to opinion diversity.

■ **Summary:**

This project demonstrates how text-level sentiment data can be transformed into meaningful, aggregated insights at the movie level. The use of randomly assigned real movie names makes the analysis more intuitive and suitable for demonstration or portfolio projects.