Title: Analyzing the Influence of Rotten Tomatoes Reviews on IMDb Movie Ratings

Author: Aditya Chunduri

Motivation/Rationale for the Project:

The project is driven by a fascination with how online reviews influence viewers' choices and movies' success. Rotten Tomatoes and IMDb are pivotal in movie selection; hence, understanding the connection between the reviews on one platform and the ratings on another can significantly impact filmmakers and audiences.

Description of Data Sources:

a. Data Sources:

1. IMDb Top 250 Movies: Extracted through web scraping to capture highly rated movies.
Link: IMDb Top 250 (https://m.imdb.com/chart/top/?ref_=nv_mv_250)

2. Rotten Tomatoes Reviews: Collected via dynamic scraping techniques, providing a comprehensive set of user reviews.
Link: Rotten Tomatoes (https://www.rottentomatoes.com/)

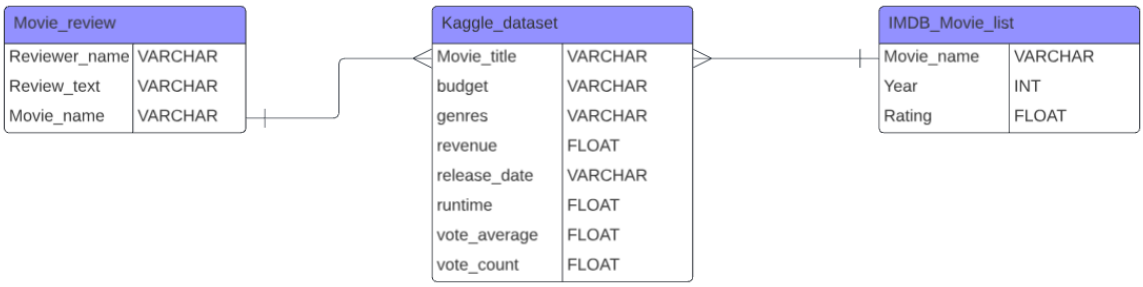
3. Kaggle Movie Dataset: Sourced from Kaggle, offering extensive details about various films.
Link: Kaggle Datasets (https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset)

b. Data Collection Changes and Challenges:

The initial plan underwent revisions to accommodate automated scraping methods due to the scope of data required. Challenges such as navigating dynamic web content and adhering to ethical scraping practices were successfully navigated.

3. **Integrated Data Model**:

The ER diagram illustrates the structure of our integrated database, combining the Movie_review, IMDB_Movie_list, and Kaggle_dataset entities. This integration is crucial for correlating Rotten Tomatoes reviews with IMDb ratings.



ER-Diagram: Depicts the data model used to integrate the collected data for analysis.

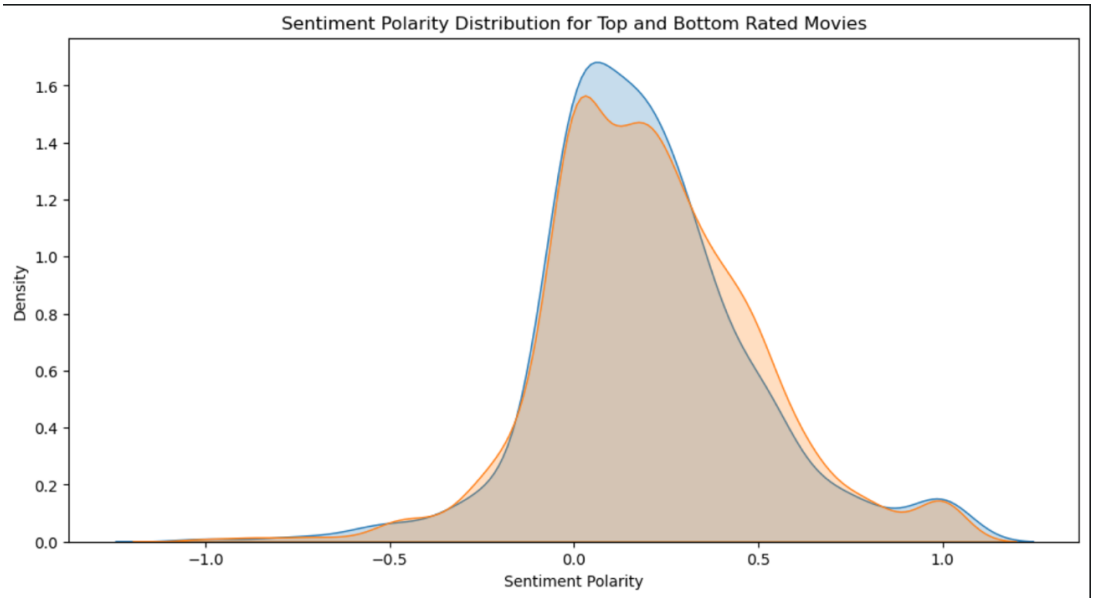
4. **Analyses/Visualizations**

a. Analysis Techniques:

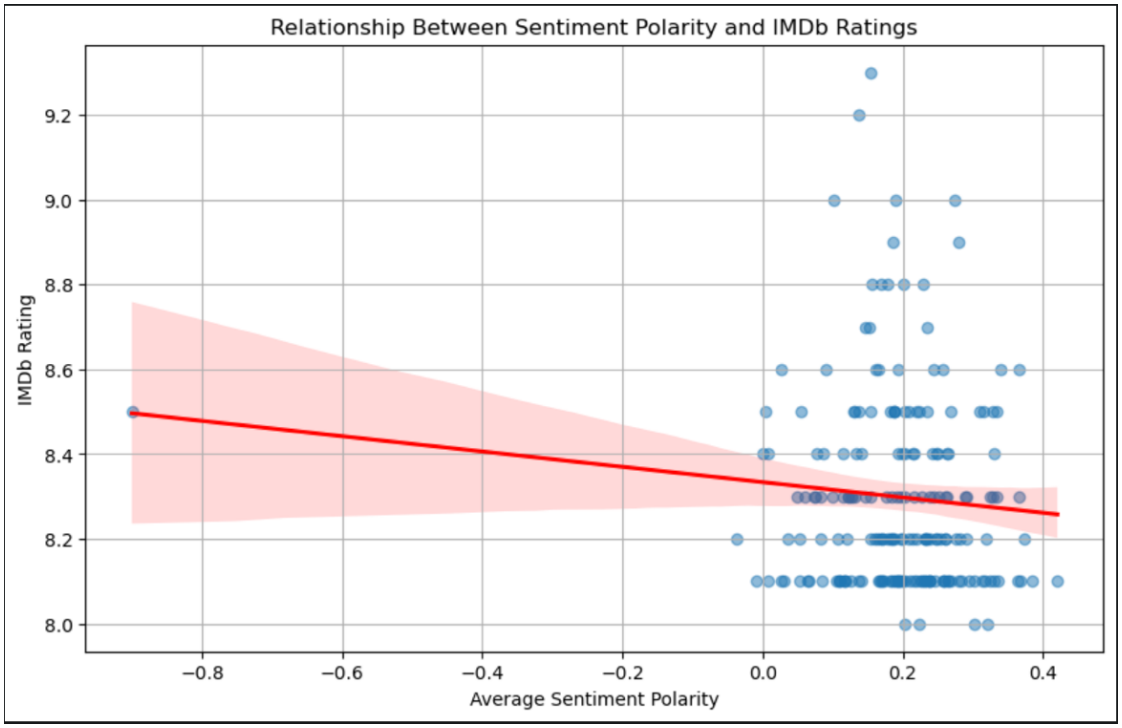
The project utilizes quantitative sentiment analysis to determine the emotional tone of Rotten Tomatoes reviews and statistical correlation analysis to assess the relationship between these sentiments and IMDb ratings.

b. Visualizations:

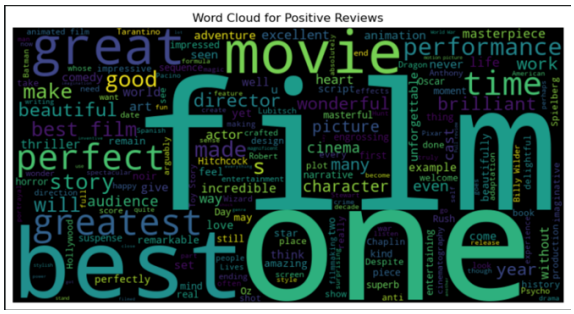
- Sentiment Polarity Distribution: Demonstrates the density of sentiment polarities for top and bottom-rated movies, indicating a skew towards positive sentiments in highly rated films.

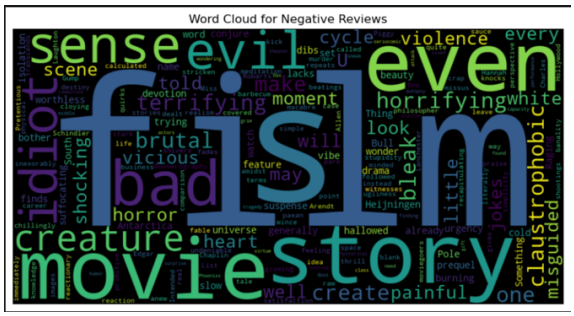


- Correlation Analysis: A scatter plot with regression analysis visualizes the relationship between sentiment polarity from Rotten Tomatoes reviews and IMDb ratings, highlighting the trend and potential correlation.

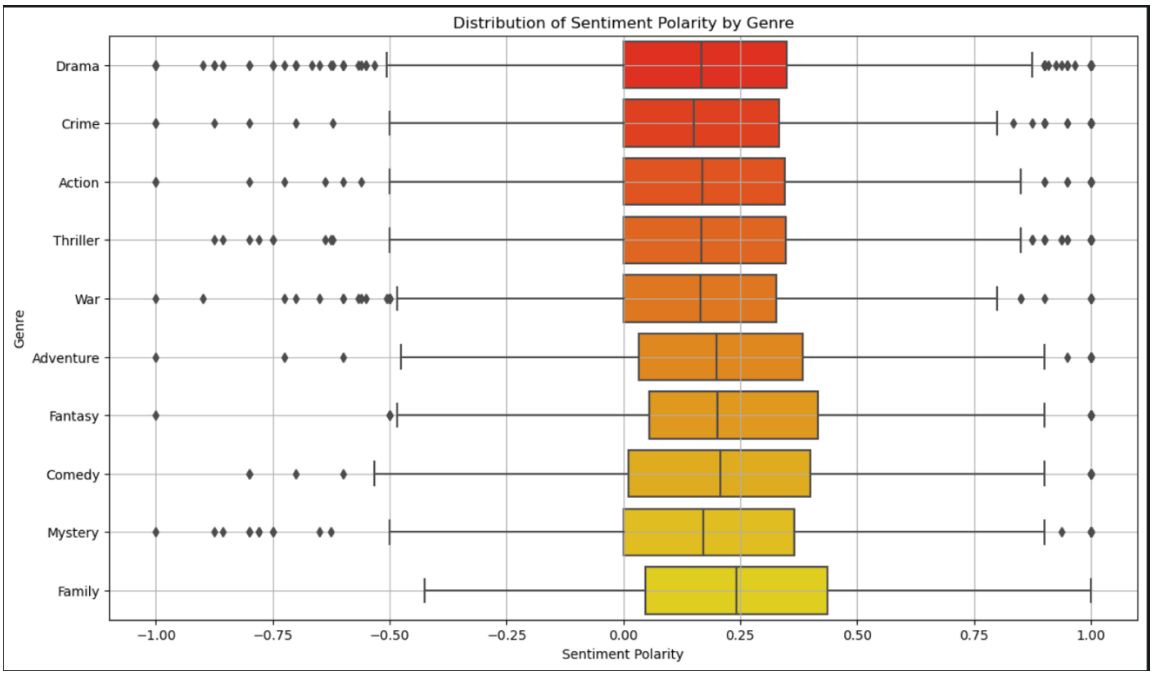


- Word Clouds for Positive and Negative Reviews: Two-word clouds displaying prevalent terms in positive and negative reviews, which helps in understanding the descriptive language influencing IMDb ratings.





- Distribution of Sentiment Polarity by Genre: This boxplot illustrates the sentiment polarity of Rotten Tomatoes reviews across different genres and investigates any patterns between these sentiments and IMDb ratings, pinpointing whether certain genres consistently receive higher ratings that correlate with their review sentiments.



Conclusions

The study indicates a modest correlation between Rotten Tomatoes reviews and IMDb ratings. It appears that films with more positive sentiment in their reviews tend to have slightly higher IMDb ratings. However, this relationship is not strongly pronounced, suggesting that while reviews do contribute to a film's rating, they are just one of several factors in a complex ecosystem influencing audience and critical reception.

Future Work

Future enhancements would include extending the dataset for robustness, exploring the impact of demographic factors, and employing predictive models to forecast IMDb ratings based on review sentiment analysis.