Trending Now: Leveraging Social Media to Identify and Analyze Movie and Genre Trends

Dimple Singh
Computer Science
Binghamton University
Binghamton, NY, USA
dsingh27@binghamton.edu

Jay Balaram Sankhe
Computer Science
Binghamton University
Binghamton, NY, USA
jsankhe1@binghamton.edu

Debangana Ghosh
Computer Science
Binghamton University
Binghamton, NY, USA
dghosh2@binghamton.edu

Jeremy Anton
Computer Science
Binghamton University
Binghamton, NY, USA
ajeremy1@binghamton.edu

Ritika Kale
Computer Science
Binghamton University
Binghamton, NY, USA
rkale2@binghamton.edu

Abstract

This project explores the potential of social media data in identifying and analyzing trends within the movie industry, specifically focusing on popular movies, TV shows and genres. By analyzing posts and comments related to movies and TV shows on platforms like Reddit and The Movie database (TMDb), we aim to uncover insights into audience preferences and emerging trends. Reddit and TMDb, two widely used platforms, generate vast amounts of text data through posts and comments. To achieve our objectives, we employed APIs from Reddit and TMDb to collect posts associated with relevant subreddits such as "r/movies" and "r/tv". We analyze the distribution of discussions and engagement across different movie genres to reveal emerging preferences and shifts in audience taste. In this analysis, we utilized various types of graphs to visually represent trends and patterns within the movie and TV genre landscape sourced from Reddit and The Movie Database (TMDb).

Collectively, our findings present a data-driven analysis of trending movies and TV genres over a specified timeframe, providing insights into the dynamic nature of the entertainment industry within this particular domain. The project demonstrates the potential of combining social media data with API-driven analyses to gain actionable insights into current trends, contributing to a more informed understanding of the evolving landscape in the movie and genre domains. This project will contribute to a deeper understanding of how social media shapes movie popularity and influences genre trends, providing valuable insights for filmmakers, distributors, and anyone interested in the dynamics of the movie industry.

Keywords: Reddit API, TMDb API, PostgreSQL, Movies, Data Collection, Analysis, Plotting, Visualization, Flask, Bootstrap, HTML, CSS.

ACM Reference Format:
Dimple Singh, Jay Balaram Sankhe, Debangana Ghosh, Jeremy Anton, and Ritika Kale. 2023. Trending Now: Leveraging Social Media to Identify and Analyze Movie and Genre Trends. In *Proceedings of ACM Binghamton conference (Binghamton'23).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/1234567890

Binghamton'23, Binghamton, NY, USA,

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/10.1145/1234567890

1 INTRODUCTION

The film and television industry is a dynamic and influential force in today's society, shaping cultural conversations and reflecting the evolving preferences of audiences. Movies and TV shows play a crucial role in communicating and expressing human experiences.

In our digital age, social media platforms have become central hubs for discussions and reflections on various forms of entertainment, including movies and TV genres. This project delves into the potential of social media data as a valuable resource for identifying and analyzing trends within the movie industry, with a specific focus on popular movies, TV shows, and genres.

We explored sentiment analysis to not just find the most popular trends but also spot instances of toxicity in discussions. Our study involved looking into the most toxic movies and shows, revealing the darker side of online conversations. Furthermore, we checked different genres to see if there are trends in toxicity linked to specific types. This detailed approach helps us give a complete picture of how people

talk about movies and TV shows online, showing that discussions can be varied and cover various aspects on social media platforms.

Our objective is to uncover insights into audience preferences and emerging trends within the movie and TV genre landscape. We've developed a dynamic site using Flask, HTML, CSS, and bootstrap to explore trending TV series, movies, and toxicity levels. Categories include "Trending TV Series" and "Movies Across TMDB" and "Reddit," as well as toxicity analysis across genres. The site aims to provide a focused and organized platform for users to delve into various aspects of entertainment trends.

2 DATASET DESCRIPTION

2.1 DATA SOURCE

This project utilizes data sourced from highly-rated movies and TV shows. TMDb rating data obtained through the TMDb API, along with posts and comments from Reddit's r/movies and r/tv via the Reddit Stream API, forms the basis of our dataset. To measure toxicity in genre-based discussions, we incorporate insights from the ModerateHatespeech API. Our implementation incorporates Flask, Python, HTML, CSS, Bootstrap and relevant libraries to construct an interactive UI. This user-friendly interface streamlines user requests and queries the stored data in PostgreSQL, enhancing the overall accessibility and functionality of our project.

2.2 API METHODS

The data collection process involves various API methods to interact with the Reddit and TMDb APIs, as well as the ModerateHatespeech API for sentiment analysis. We utilize HTTPS requests, primarily employing GET and POST methods.

Our Flask web application features the following API routes:

- GET api/tmdb_movies: Results for top trending movies from TMDB.
- GET api/tmdb_tv: Results for top trending TV shows from TMDB.
- GET api/movies_list: Results for top trending movies from Reddit.
- GET api/tv_list: Results for top trending TV shows from Reddit.
- GET api/genre_movies: Results for top trending genres of movies like comedy, drama, action with most toxic discussions.
- GET api/genre_tv: Results for top trending genres for TV shows like drama, sci-fi, and adventure with most toxic discussions.
- GET api/tv_toxic_list: Results with names of TV shows with the most toxic discussions.
- GET api/movies_toxic_list: Results with names of movies with the most toxic discussions.

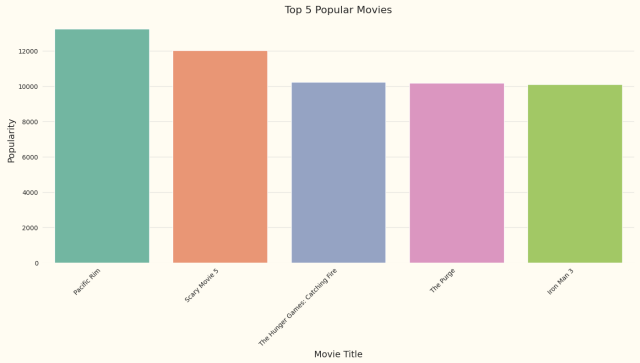


Figure 1. Trending Top Five Movies

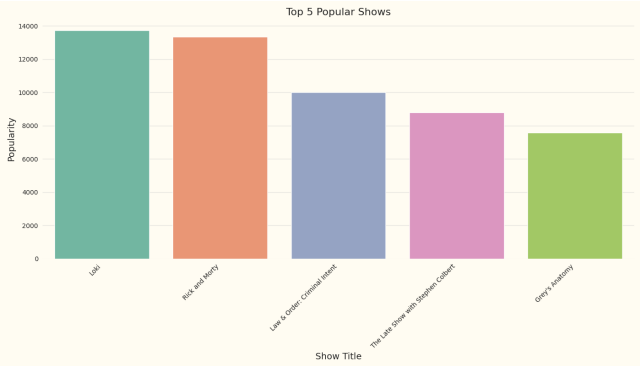


Figure 2. Trending Top Five TV Shows

The Figure 1 showcases the top trending movies, the five most popular films across the social media platforms and discussions. From Figure 1, we can see *Pacific Rim* is the most popular movie followed by S*cary Movie 5*, *The Hunger Games*, *The Purge* and *Iron Man 3*. Similarly, Figure 2 provides insights into the trending TV shows, highlighting the current favorites among audiences. From the Figure 2 We can see Loki is the most popular TV show.

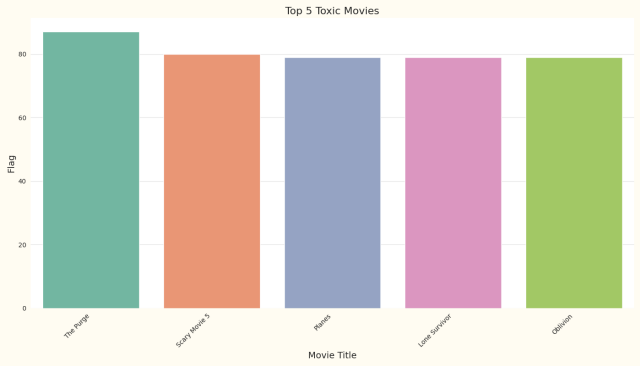


Figure 3. Top Five Toxic Movies

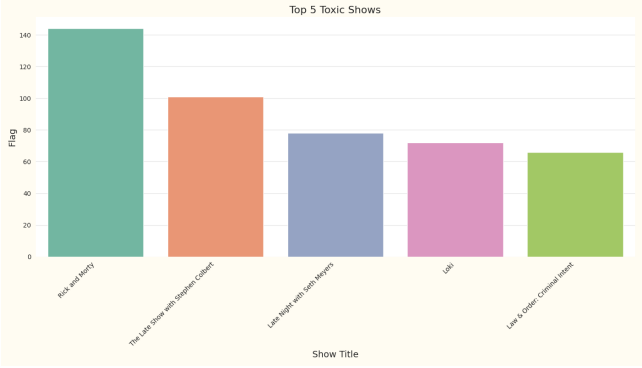


Figure 4. Top Five Toxic TV Shows

Figures 3 and 4 delve into toxicity within the entertainment sphere, revealing the top five movies and TV shows associated with discussions containing toxic elements. *The Purge* and *Rick and Morty* are most toxic Movies and TV shows.

3 BACKGROUND AND ANALYSIS

As social media increasingly shapes public discourse and trends, there is a rising interest in harnessing platforms like Reddit to gain insights into discussions surrounding popular movies and TV shows. These platforms serve as valuable sources of real-time data, offering a glimpse into audience sentiments and preferences.

Our primary focus is on displaying are results on the api for the comprehensive collection and analysis of discussions related to currently trending movies and TV shows. Acknowledging the significant impact of online discussions on the perception and success of entertainment content, our research aims to bridge the gap between social media chatter and quantifiable insights into entertainment trends.

A key development in our project is the integration of our database with aur website (APIs). ModerateHatespeech API, used for real-time toxicity measurement. This API enables us to conduct sentiment analysis on discussions about movies and TV shows, classifying them as either 'normal' or 'flagged' for toxicity with a confidence score, providing a nuanced understanding of online conversations.

The project encountered various challenges, particularly in obtaining timely data, initially facing issues with limited data from Reddit. However, overcoming these challenges highlighted the dynamic nature of social media data, showcasing the project's problem-solving capacity and innovation.

While integrating the ModerateHateSpeech API, instances of empty responses were addressed by refining the script to discard these responses and handle potential exceptions. Time constraints led to incomplete processing of the entire data body, emphasizing the need for additional time to gather more comprehensive insights into toxicity across genres and titles.

Ethical considerations in handling user-generated content and sentiment analysis are paramount, recognizing the importance of respecting user privacy and ensuring responsible use of collected data.

In summary, our work extends previous research in social media analytics and sentiment analysis. By combining data from Reddit discussions, TMDb ratings, and toxicity measurements, we aim to provide valuable insights into the evolving landscape of entertainment trends and audience preferences.

4 DATA EXPLORATION

When exploring the sentiment within social media discussions, we specifically focused on analyzing the distribution of toxicity in movie and TV show discussions, emphasizing the genre-specific aspects based on the data gathered from Reddit. Toxicity Analysis of Television Genres: As illustrated in Figure 4 , our examination identified "*Rick and Morty*" as a central topic in television discussions, with 414 instances flagged as toxic in comparison to other TV shows. Toxicity Analysis of Movie Genres: In the context of movies, as depicted in Figure 3, "The Purge" demonstrated 87 instances flagged as toxic compared to other movies.

5 IMPLEMENTATION AND DASHBOARD

For the dashboard, we made a web-app using Python flask, CSS and HTML, which takes number of titles as input and plots the requested plot between number of flagged comments and the title name for both movies and tv shows. When we execute the flask python script, as seen in Figure 5, the web-app is hosted on http://128.226.29.118 with apache server working as a reverse proxy, and we can input the number of titles and generate plots there. When we input the number and click the generate plot button, in the backend, a function is called with the number as input parameters, and the function returns a list of the titles that are most toxic in descending order and return the data to the flask app, which generates the plot for the same using matplotlib, base64 and seaborn. We accessed the VM to make the dashboard and generate the URL. Also note that we had to be on our university VPN to access the VM.

As we can see in the Figure 5, it represents the home page of our flask web-app,where we can enter the number to titles the user want to compare and generate the plots for. Figure 6 show us the top trending titles which are most toxic for tv shows and movies under separate links.

Figure 5. Home Page of Website

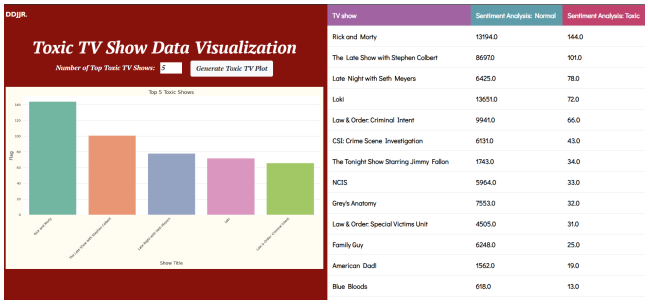
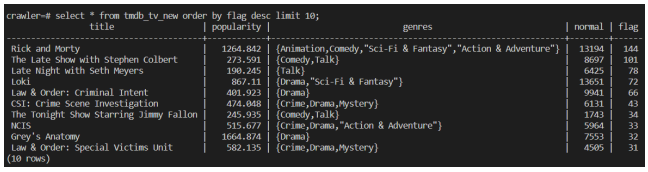


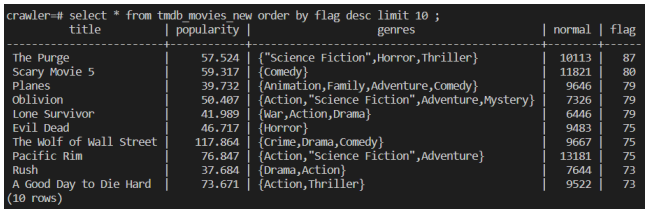
Figure 6. Top Trending Titles for Toxic TV Shows



```
crawler=# select * from tmdb_tv_new order by flag desc limit 10;
             title              | popularity |                  genres                  | normal | flag
--------------------------------+------------+------------------------------------------+--------+------
 Rick and Morty                 |   1264.842 | {Animation,Comedy,"Sci-Fi & Fantasy","Action & Adventure"} |  13194 |  144
 The Late Show with Stephen Colbert |   273.591 | {Comedy,Talk}                            |   8697 |  101
 Late Night with Seth Meyers    |   190.245  | {Talk}                                   |   6425 |   78
 Loki                           |    867.11  | {Drama,"Sci-Fi & Fantasy"}               |  13651 |   72
 Law & Order: Criminal Intent   |   401.923  | {Drama}                                  |   9941 |   66
 CSI: Crime Scene Investigation |   474.048  | {Crime,Drama,Mystery}                    |   6131 |   43
 The Tonight Show Starring Jimmy Fallon |   245.935 | {Comedy,Talk}                    |   1743 |   34
 NCIS                           |   515.677  | {Crime,Drama,"Action & Adventure"}       |   5964 |   33
 Grey's Anatomy                 |  1664.874  | {Drama}                                  |   7553 |   32
 Law & Order: Special Victims Unit |   582.135 | {Crime,Drama,Mystery}                    |   4505 |   31
(10 rows)
```

Figure 7. TV Shows by Toxicity in Database

Figure 7 encompasses a compilation of the ten TV shows organized by their toxicity on TMDb, stored within the database. Each show's title is accompanied by relevant details such as its popularity score, genres, the count of normal votes, and a flag value. Notably, the TV show ranked as the most toxic on the list is "Rick and Morty," holding a flag score of 144. Conversely, the tenth position is occupied by "Law and Order: Special Victims Unit," with a flag score of 31.



```
crawler=# select * from tmdb_movies_new order by flag desc limit 10 ;
          title           | popularity |                  genres                  | normal | flag
--------------------------+------------+------------------------------------------+--------+------
 The Purge                |     57.524 | {"Science Fiction",Horror,Thriller}      |  10113 |   87
 Scary Movie 5            |     59.317 | {Comedy}                                 |  11821 |   80
 Planes                   |     39.732 | {Animation,Family,Adventure,Comedy}      |   9646 |   79
 Oblivion                 |     50.407 | {Action,"Science Fiction",Adventure,Mystery} |   7326 |   79
 Lone Survivor            |     41.989 | {War,Action,Drama}                       |   6446 |   79
 Evil Dead                |     46.717 | {Horror}                                 |   9483 |   75
 The Wolf of Wall Street  |    117.864 | {Crime,Drama,Comedy}                     |   9667 |   75
 Pacific Rim              |     76.847 | {Action,"Science Fiction",Adventure}     |  13181 |   75
 Rush                     |     37.684 | {Drama,Action}                           |   7644 |   73
 A Good Day to Die Hard   |     73.671 | {Action,Thriller}                        |   9522 |   73
(10 rows)
```

Figure 8. Movies by Toxicity in Database

In Figure 8 list of 10 movies ordered by their toxicity on TMDb is stored within the database.The most popular movie on the list is *"The Purge"* with a flag score of 87, and on number 10 is *"A Good Day to Die Hard"* with a flag score of 73.

6 DATA FLOW

Our project methodology centers on two crucial components: the methodical collection of data from diverse sources and the processing of this data for sentiment analysis. This approach aims to uncover the nuances within online discussions surrounding movies and TV shows, specifically focusing on their sentiment implications. Utilizing our Flask App API, we effectively handle user queries, triggering interactions with the database to retrieve pertinent information. The application then seamlessly delivers responses to the webpage based on user inquiries. This systematic process guarantees a dynamic and responsive user experience for those seeking insights into the sentiment dynamics of media discussions. The entire process is illustrated in the flow chart presented in Figure 9.

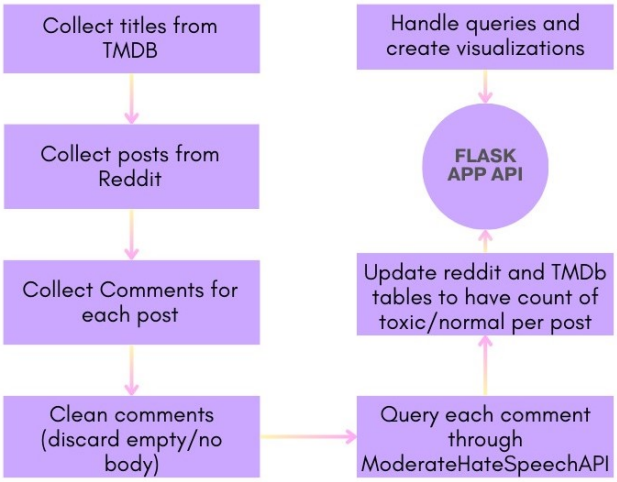


Figure 9. Flow Diagram

7 OBJECTIVES

The primary goals of our project involve analyzing and comprehending the social media discussions pertaining to movies and TV shows, with a specific focus on investigating the sentiment and toxicity across various genres and titles of movies and tv shows. These objectives aim to contribute to both academic research and practical applications in media analysis and trend forecasting. Furthermore, our attention is directed towards addressing the following research questions to further expand our work:

Does the popularity of movies or television correlate with the amount of toxic content on online forums? Additionally, which genres generate the highest levels of toxic content?

8 RESULTS

Our exploration into social media data, focusing on movies, TV shows, and genres, has yielded valuable insights. Through the integration of TMDb rating data, Reddit posts, and sentiment analysis using the ModerateHatespeech API, we have constructed a robust dataset. The user-friendly Flask web application, coupled with PostgreSQL, provides an interactive interface for seamless data retrieval and analysis.

The top trends in our data exploration reveal interesting patterns. For TV series and movies across TMDB, the top picks reflect the current favorites among users. On Reddit, we identified the trending TV series and movies based on user discussions, offering a unique perspective shaped by community engagement.

Shifting focus to toxicity analysis, we uncovered the levels of toxicity in TV series and movies, highlighting content sparking intense online discussions. Expanding our view, we explored toxicity across TV and movie genres, revealing patterns in genre-based discussions. These results showcase the diverse landscape of trends and toxicity within the realm of TV series and movies across both TMDB and Reddit platforms.

9 CONCLUSION

In conclusion, this project has successfully demonstrated the potential of social media data, API-driven analyses, and sentiment analysis in uncovering trends within the movie industry. By focusing on popular content, genre preferences, and toxicity levels, we've created a user-friendly platform that contributes to a deeper understanding of audience engagement on social media platforms. The project's dynamic site showcases the multifaceted aspects of entertainment trends, emphasizing the importance of considering both positive trends and potential toxicity in audience interactions. Overall, this work contributes to academic research and lays the groundwork for continued exploration into the evolving impact of social media on entertainment trends.

10 ACKNOWLEDGEMENT

We would like to acknowledge the Reddit and TMDB APIs for providing access to valuable data. We also appreciate the guidance and support received during the project's development.

References

[1] Meet Singh, Dhrubasish Sarkar. *Content-Based Movie Recommendation System with Sentiment Evaluation of Viewer's Reviews.* https://link.springer.com/chapter/10.1007/978-981-16-6893-7_15

[2] Sandipan Sahu, Raghvendra Kumar, Mohd Shafi Pathan, Jana Shafi, Yogesh Kumar, Muhammad Fazal Ijaz. *Movie Popularity and Target Audience Prediction Using the Content-Based Recommender System.* https://ieeexplore.ieee.org/abstract/document/9758691