

# Analysis of the 2022 Anime Database

STOR 520 Group 2

QI XUE  
MINLI ZHENG  
LEI ZHANG  
JIAYI CHEN



# Contents

**01. Introduction**

**02. Research Questions**

**03. Data Source**

**04. Sentiment Analysis**

**05. Predictive Models**

**06. Conclusion**

01

# Introduction



# Introduction



## Why choose anime for analysis:

- Cultural Impact: Anime is a significant global cultural phenomenon, offering a window into both Japanese and international cultures.
- Popularity Among Youth: With a strong following among teenagers, studying anime helps understand this key demographic's preferences.
- Insights for Creators: The findings can inform anime creators and distributors about audience preferences and trends.



## Target user:

The primary audience for this study includes creators, distributors, and enthusiasts who are keen on understanding the emotional trends and audience preferences in shoujo and shounen anime.



## Focus data:

- Shoujo
- Shounen

02

# Research Questions



# Research Questions

---

- How can sentiment analysis techniques be used to understand the emotional trends in shoujo and shounen anime over the years?
- When examining the scores of anime targeting teenage audiences, which models are used to ensure the better predictions?



03

# Data Source



# Data Source

Show 3 entries

Search:

	ID	Title	Synopsis	Type	Episodes	Start_Aired	Source	Genres	Themes
1	16498	Shingeki no Kyojin	Centuries ago, mankind...	TV	25	7-Apr-13	Manga	Action, Drama	Gore, Militar Survival
2	1535	Death Note	Brutal murders, petty...	TV	37	4-Oct-06	Manga	Supernatural, Suspense	Psychologic
3	5114	Fullmetal Alchemist: Brotherhood	After a horrific alchemy...	TV	64	5-Apr-09	Manga	Action, Adventure, Drama, Fantasy	Military

Showing 1 to 3 of 2,701 entries

Previous

1

2

3

4

5

...

901

Next

- **Source:** The "Anime Database 2022" from Kaggle, originally compiled through web scraping from MyAnimeList.net.
- **Volume of Data:** The dataset contains details of 21,460 anime titles.
- **Dataset Size:** Our targeted dataset includes 2,701 entries specific to shoujo and shounen anime.
- **Key Variables:** Each entry includes information like Title, Favorites, Ratings, and Members. For deeper analysis, we also considered variables like Themes and Genres.

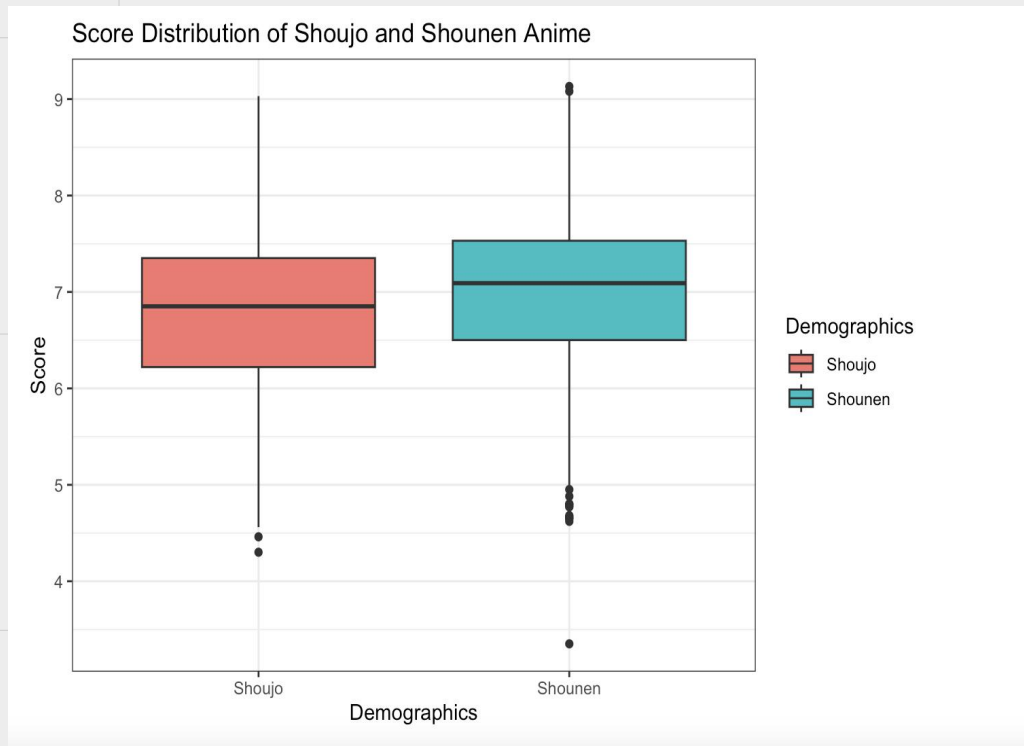


# Data Source

*We aim to analyze the distribution of scores for shoujo and shounen anime to aid in the development of predictive models...*

**Median Scores:** Both shoujo and shounen anime genres have a median score around 7.

**Outliers:** Shounen anime exhibits more outliers, showing greater variability in scores, including some significantly low scores.



04

# Sentiment Analysis

(Shoujo and Shounen)



# Sentiment Analysis—Introduction



## Overview of Shoujo and Shounen Anime

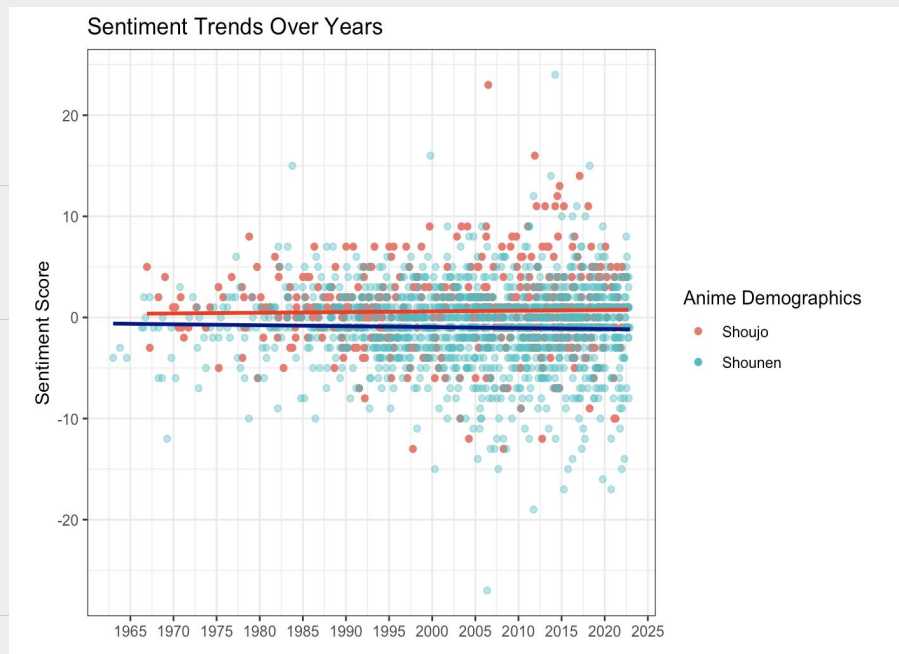
- **Shoujo Anime:** Translating to "girls" in Japanese, shoujo anime targets young girls and women. These series often explore themes of **relationships, emotional growth, love**, and the challenges of **adolescence**.
- **Shounen Anime:** Meaning "boys" in Japanese, shounen anime is geared towards young boys and men, typically **aged 12-18**. These series focus on action, adventure, camaraderie, and often incorporate themes of **bravery, strength, and overcoming adversity**.



## Sentiment Analysis Method

- **Data Selection:** We used the 'Demographics' variable from a comprehensive anime dataset with over **20,000 entries** to focus on shoujo and shounen anime.
- **Analysis Process:** Applying the **tidytext and Bing library**, we performed an inner join with the Bing lexicon, categorizing words into positive (+1) and negative (-1) groups, to calculate overall sentiment scores from the **'Synopsis'** variable across years.

# Sentiment Analysis——1966-2022 Sentiment Trends



## ■ Sentiment Trends

During this period, both shoujo and shounen anime exhibited relatively *stable* sentiment scores. Shoujo anime tended to have *higher positive* sentiment scores, suggesting that these works contain more elements of happiness, joy, and love. In contrast, shounen anime showed more elements of sadness, anger, and violence.

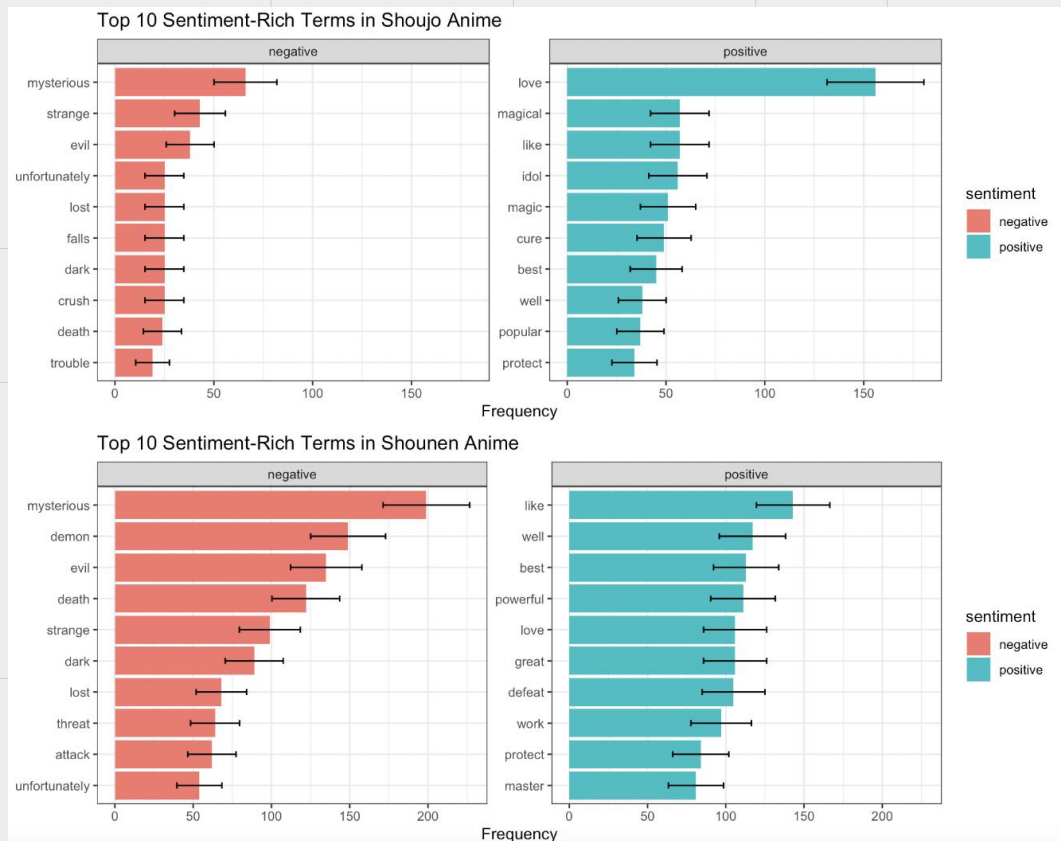
## ■ Specific Year Negative Sentiment Peak

*"Pretty Guardian Sailor Moon Eternal Part 2"*  
(Released in 2021)

This anime portrays a post-solar eclipse Earth shrouded in darkness, presenting a potentially *severe challenge* for the characters, thus eliciting a strong negative emotional response.

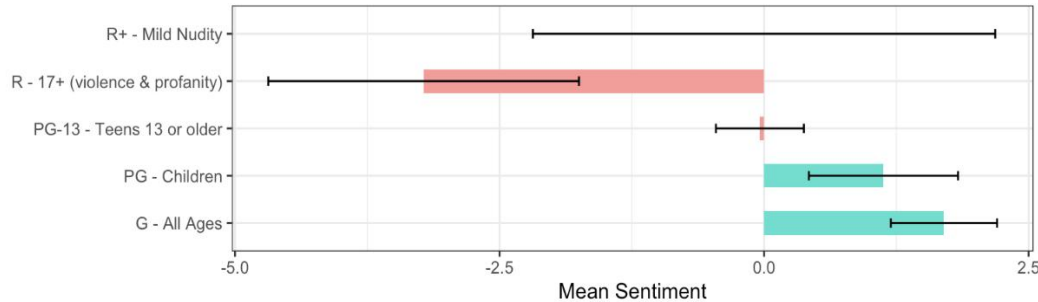
# Sentiment Analysis—Positive and Negative Themes

- **Shoujo Anime:** "Love" is the most frequent term, emphasizing romance as a key theme. Other common terms like "magical" suggest a blend of romance and fantasy preferred by young female audiences.
- **Shounen Anime:** Terms like **personal growth, power, and victory** are prevalent, reflecting themes of strength and overcoming challenges typical in narratives for young male audiences.
- **Both Genres:** The term "mysterious" appears frequently as a negative term, indicating its role in adding intrigue and complexity to the storytelling in both shoujo and shounen anime.

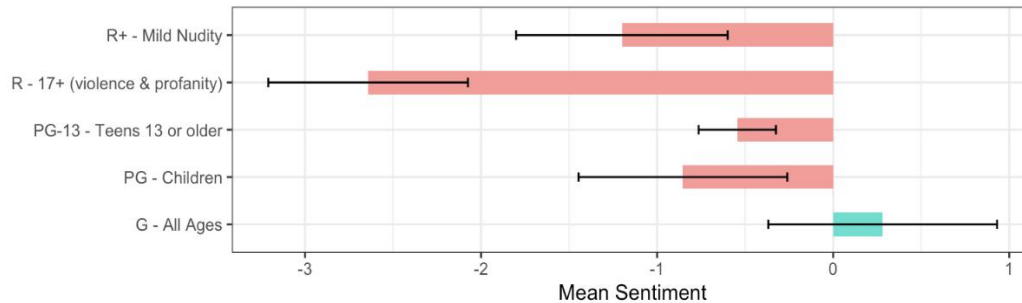


# Sentiment Analysis—Impact of Ratings on Sentiment Scores

Sentiment by Ratings in Shoujo Anime



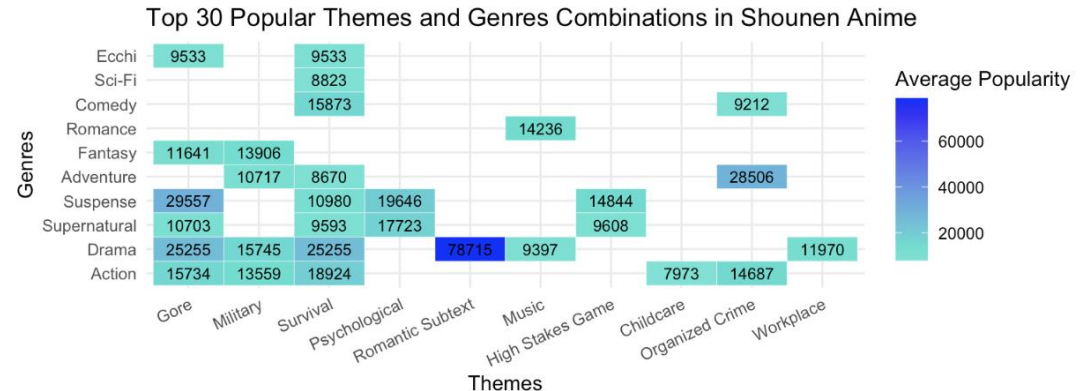
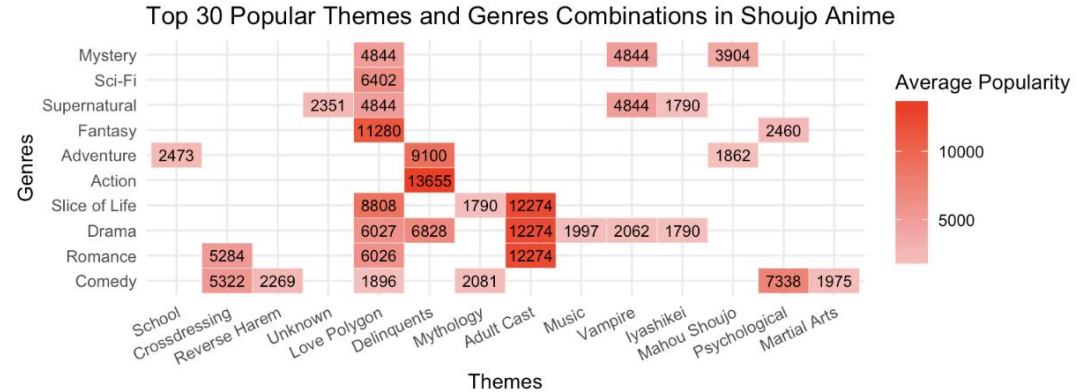
Sentiment by Ratings in Shounen Anime



- **Low Sentiment in R-Rated Anime:** "R - 17+ (violence & profanity)" rated anime shows the lowest sentiment scores in both shoujo and shounen, likely due to their **violent and profane** content.
- **Unexpected PG-13 Results:** Surprisingly, "PG-13 - Teens 13" rated shounen anime also registered negative sentiment scores, contrary to expectations of more positive content.

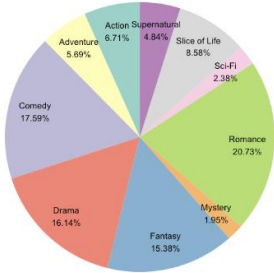
© 2006 The Authors  
Journal compilation © 2006 Blackwell Publishing Ltd

- Preferred combinations: "Suspense" and "Gore", and "Adventure" and "Organized Crime" for thrilling stories.
- Surprisingly, "Drama" and "Romantic Subtext" is the most favored, indicating a liking for emotional depth and subtle romance, despite the action-oriented reputation of shounen anime.

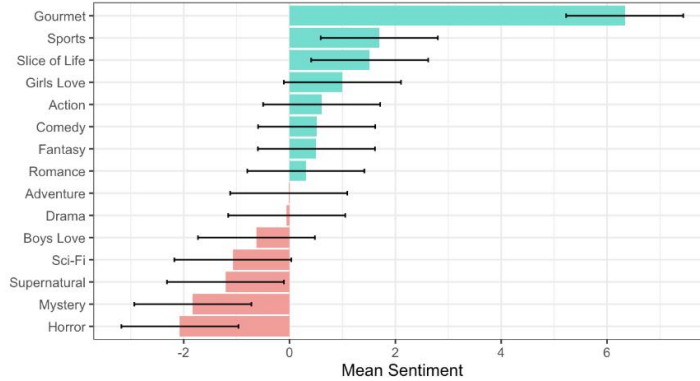


# Sentiment Analysis—Genre Sentiment

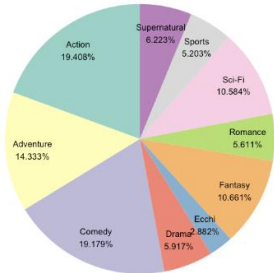
Top 10 Genres in Shoujo Anime



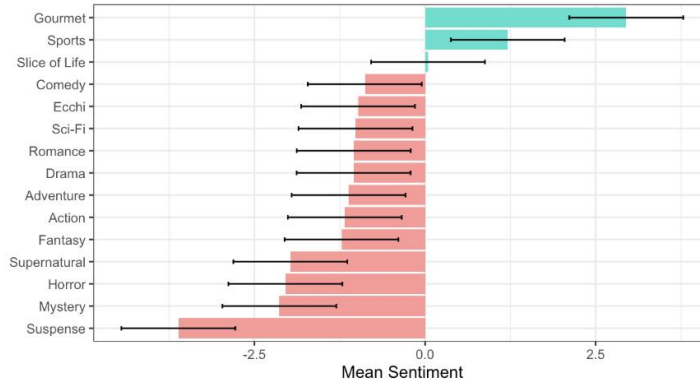
Mean Sentiment by Genre in Shoujo Anime



Top 10 Genres in Shounen Anime



Mean Sentiment by Genre in Shounen Anime



## Shounen Anime:

- Generally negative sentiment due to themes of **conflict** and **violence**.
- Themes like "death", "fight", "kill", "betrayal", and "suffering" contribute to lower sentiment.

## Shoujo Anime:

- Diverse genres, with **romance** being prominent.
- Fantasy genre indicates a desire for escapism and imaginative worlds.

## Boys Love (BL) vs. Girls Love (GL):

- BL anime tends to have a **negative** sentiment due to complex relationships and explicit themes.
- GL anime emphasizes emotional connections, resulting in a more **positive** sentiment score.
- BL anime often carries an "R - 17+ (violence & profanity)" rating, contributing to its sentiment score.



05

# Predictive Models



# Predictive Models—Introduction

*In the process of developing a regression model to predict anime scores for adolescent viewers (Shounen and Shoujo demographics), the following key steps were taken...*

## Data Preparation

- Removed extraneous columns
- Removed rows with null values, resulting in a dataset of 2449 entries out of 2771.

## Data Standardization

Standardized the numerical data to ensure consistent scale

## Model Selection

Chose three models: Linear Regression, Random Forest Regressor, and XGBoost Regressor.

## Cross-Validation

Employed a 5-fold cross-validation technique

## Model Evaluation

Assessed and compared the models' performance using RMSE

# Predictive Models—Model Evaluation

Show  entries

Model	Parameters	Best.RMSE
Null Model	Null	0.11276
Linear Regression	Null	0.10069
Random Forest	max_depth: 30, min_samples_split: 5, n_estimators: 500	0.07184
XGBoost	learning_rate: 0.1, max_depth: 3, n_estimators: 200	0.07156

Showing 1 to 4 of 4 entries

Previous  Next

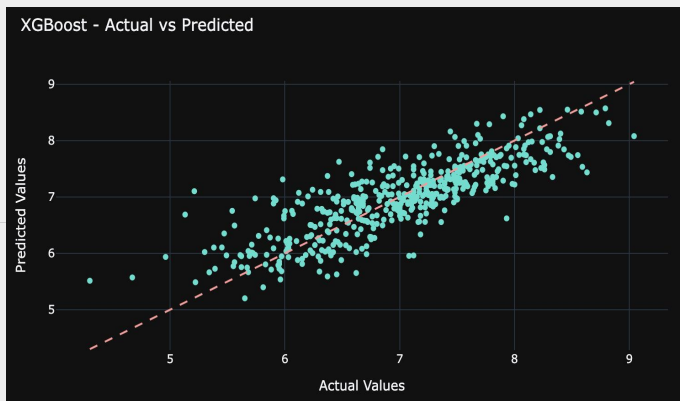
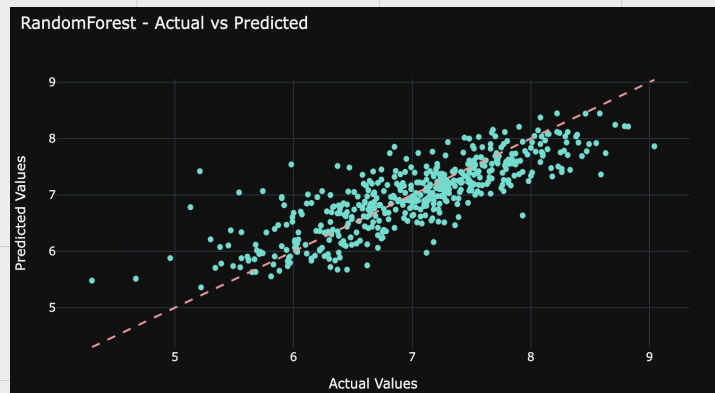
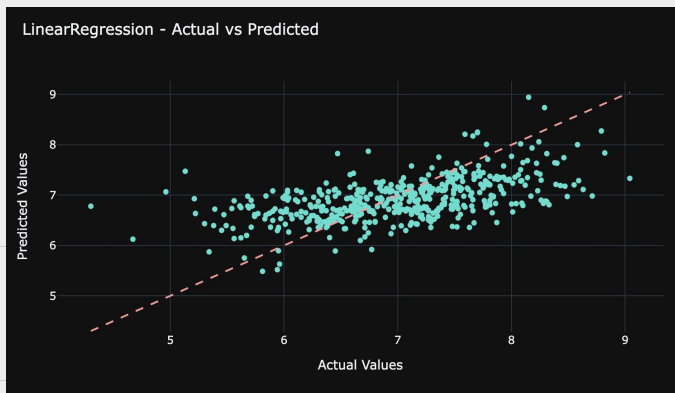
- Both Random Forest and XGBoost models outperform the baseline linear regression model.
- XGBoost leads by a narrow margin, indicating superior predictive capabilities and robustness.

# Predictive Models——Feature Impact in Linear Regression Model

$$\begin{aligned}\hat{\text{Score}} = & 1.8584 \\ & + (0.0229 \times \text{Type}) \\ & + (0.0159 \times \text{Episodes}) \\ & + (0.0204 \times \text{Source}) \\ & + (-0.0034 \times \text{Demographics}) \\ & + (-0.0005 \times \text{Duration\_Minutes}) \\ & + (-0.1134 \times \text{Rating}) \\ & + (-0.0505 \times \text{Scored\_Users}) \\ & + (-0.0950 \times \text{Members}) \\ & + (-0.2069 \times \text{Favorites})\end{aligned}$$

- Key variables impacting the Score prediction: **Favorites, Rating, and Members.**
- These variables have substantial coefficients, indicating significant influence on Score predictions.
- **Duration\_Minutes** has a minimal effect on Score prediction (coefficient of 0.0005).

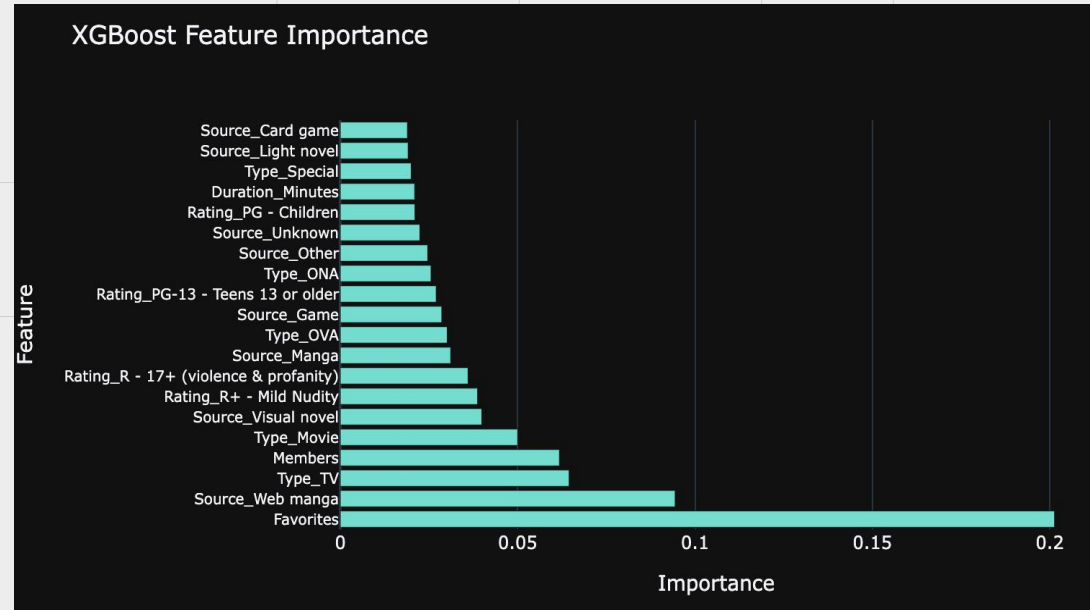
# Predictive Models—Model Performance with Scatter Plots



- Linear Regression has more **spread**, indicating limitations in capturing trends.
- Random Forest shows **tighter** clustering, implying better fit and accuracy.
- XGBoost exhibits a **similar** pattern to Random Forest with robust predictions.

# Predictive Models—Key Factors in Predicting Anime Scores

- Favorites feature is the most significant predictor, followed by Members, Source\_Manga, and Scored\_Users.
- Community engagement metrics ((Favorites, Members, Scored\_Users)) have the highest impact on anime scores, followed by content-related features (Source, Type, Rating), with nuances based on demographics and source type.



# Predictive Models— Cross-Platform Analysis

Show  entries

Title	Demographics	Predicted_Score	IMDb_Rating
Akuyaku Reijou nanode Last Boss wo Kattemimashita	Shoujo	7.18189	6.9
Bishoujo Senshi Sailor Moon Cosmos Movie	Shoujo	6.886652	7.7
Detective Conan: Love Story at Police Headquarters - Wedding Eve	Shounen	6.260386	6.3

Showing 1 to 3 of 3 entries

Previous  Next

- Supported the reliability of predictions by analyzing [audience comments](#).
- [High consistency](#) between model predictions and IMDb scores, but comparisons should be considered relative due to differences in scoring systems and user communities.
- Suggests comparing [percentile rankings](#) of the same anime on both platforms for a more accurate assessment of the model's precision.

**06**

# Conclusion





# Conclusion

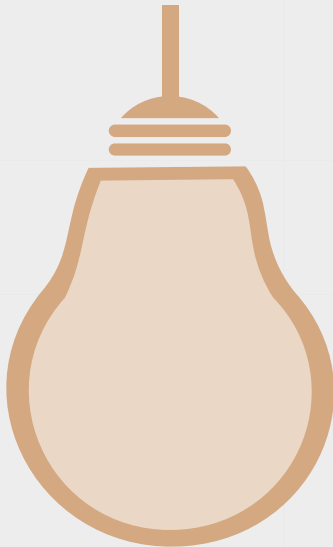
---

## Predictive Modeling

- ✓ Employed advanced machine learning (XGBoost) to predict anime scores for teens.
- ✓ Showcased the importance of community engagement and content characteristics in anime success

## Themes and Genres

- ✓ Highlighted romance in shoujo and personal growth in shounen anime.
- ✓ Emphasized a preference for non-traditional narratives.



## Sentiment Analysis

- ✓ Revealed consistent emotional tones in shoujo and shounen anime.
- ✓ Noted occasional spikes in negative sentiment.

## Sentiment in Themes

- ✓ Explored the nuanced sentiment differences, such as "Boys Love" and "Girls Love".
- ✓ Revealed insights into the emotional complexity of anime storytelling.

A soft, circular watercolor splash in shades of light green and teal serves as a background for the text. The splash has a textured, painterly appearance with varying intensities of green.

thankyou