

COD

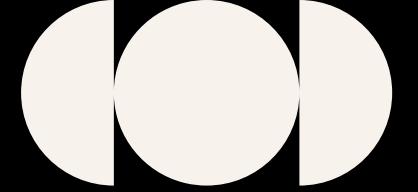
# ANIME RECOMMENDER SYSTEM



PRESENTED BY  
TEAM EG3



DOD



# Team



**JUSTIN  
TSHIFARO**



**LUNGILE  
BALOYI**



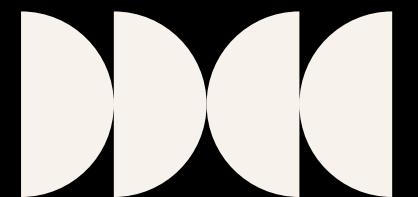
**ZWIITWAHO  
MUGODO**

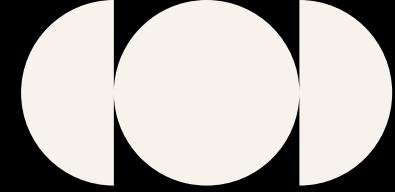


**MAHLATSE  
MASEMOLA**



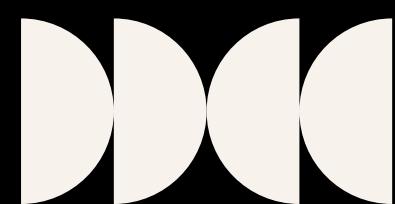
**MBALENHLE  
LENEPA**





# TABLE OF CONTENT

- 1. Introduction
  - 2. Data Cleaning
  - 4. Exploratory Data Analysis (EDA)
  - 5. Models
  - 6. Preprocessing
  - 7. Streamlit
  - 8. Recommendations
  - 9. Conclusion
- 



COD

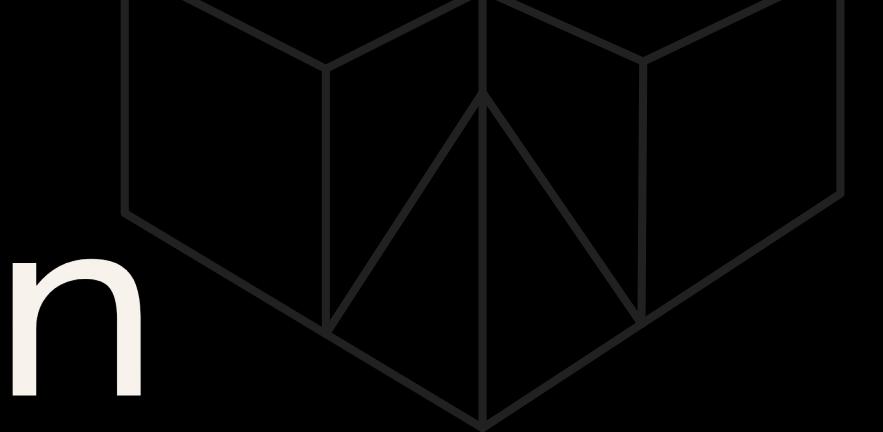
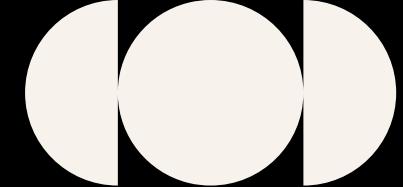
# INTRODUCTION

## Background:

- Anime growth leads to overwhelming content.
- Traditional recommendations lack accuracy and personalization.
- Hard to find preferred series.

## Problem Statement:



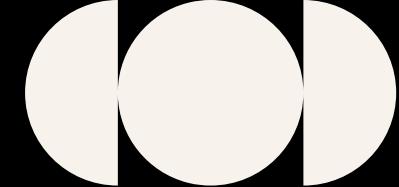


# Introduction

## About the Dataset:

- **Source:** Kaggle Anime Recommender System Project 2024 competition.
- **Content:**
  - user info,
  - anime names,
  - ratings
  - genres
  - types
  - number of episodes and members.





# Introduction



## Project Aims:

1

Develop advanced anime recommendation system.

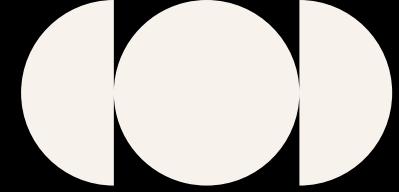
2

Provide accurate, personalized suggestions.

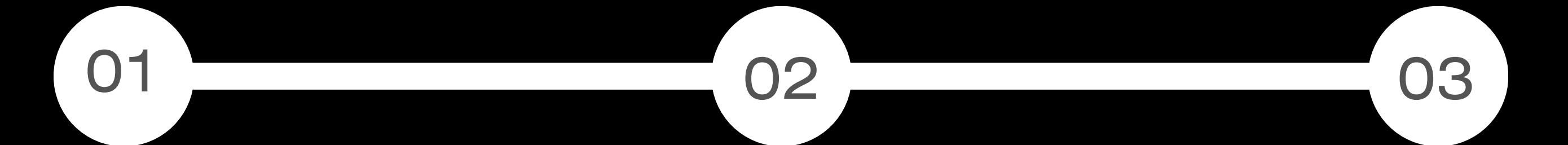
3

Reduce search time and boost engagement.





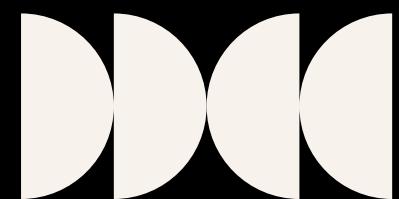
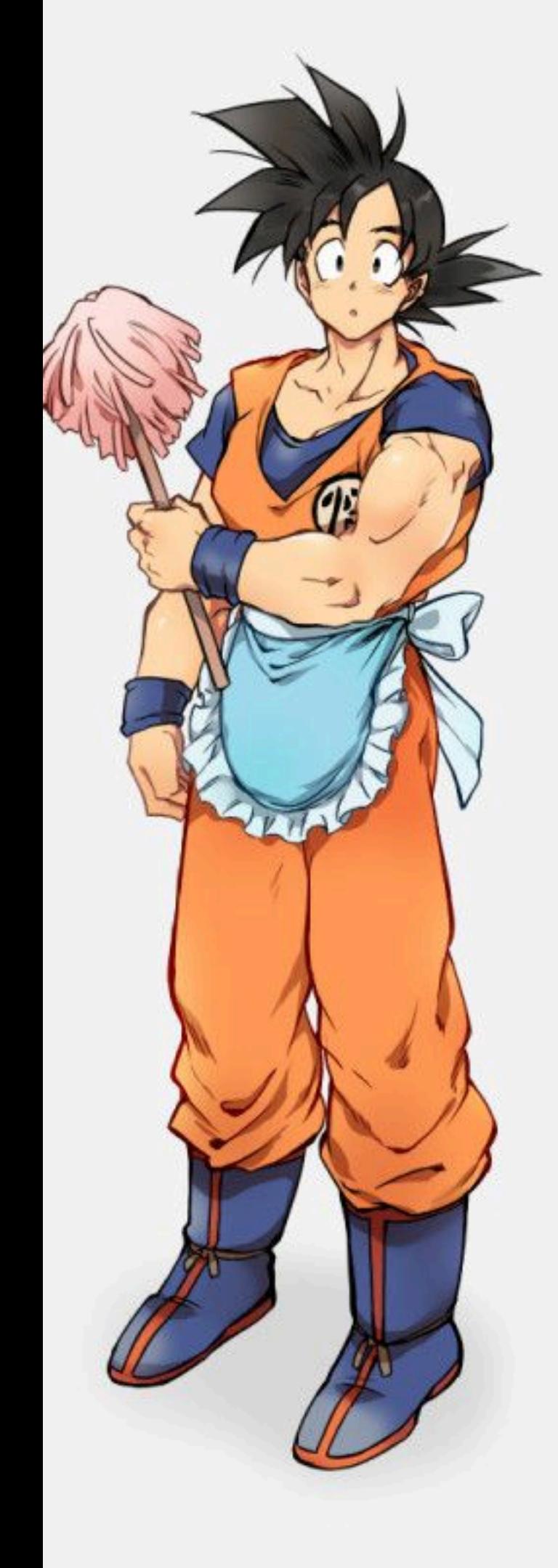
# Data Cleaning

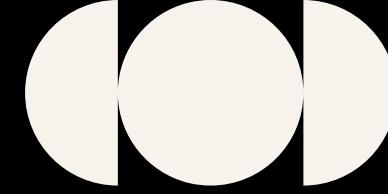


Filled in missing values in data using averages or most common values.

Removed duplicate to ensure each record is unique.

Converted data types where necessary for accurate analysis.

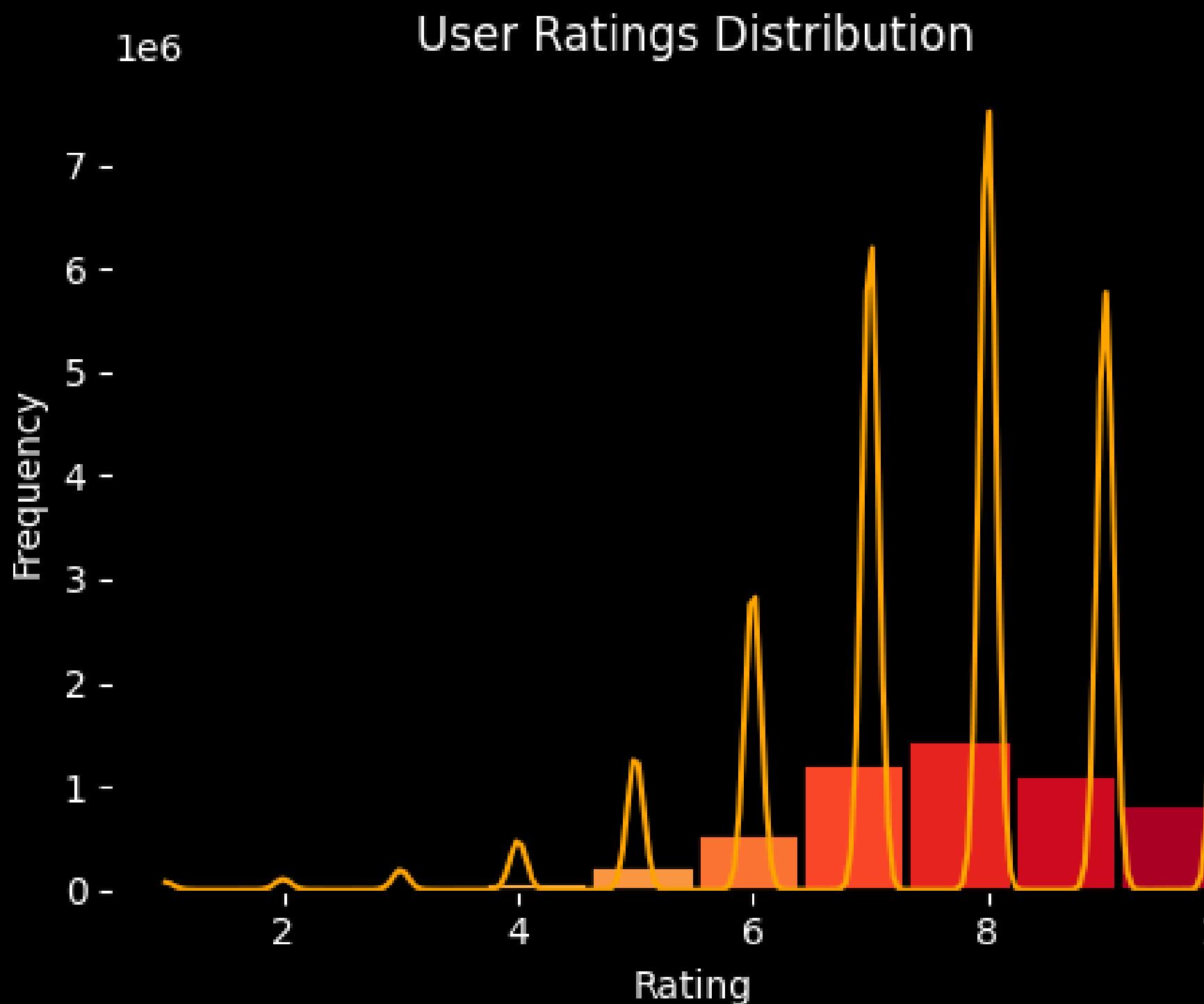




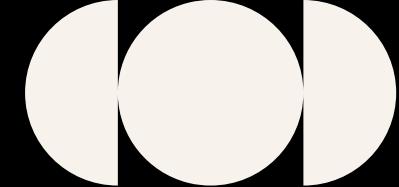
# Exploratory Data Analysis



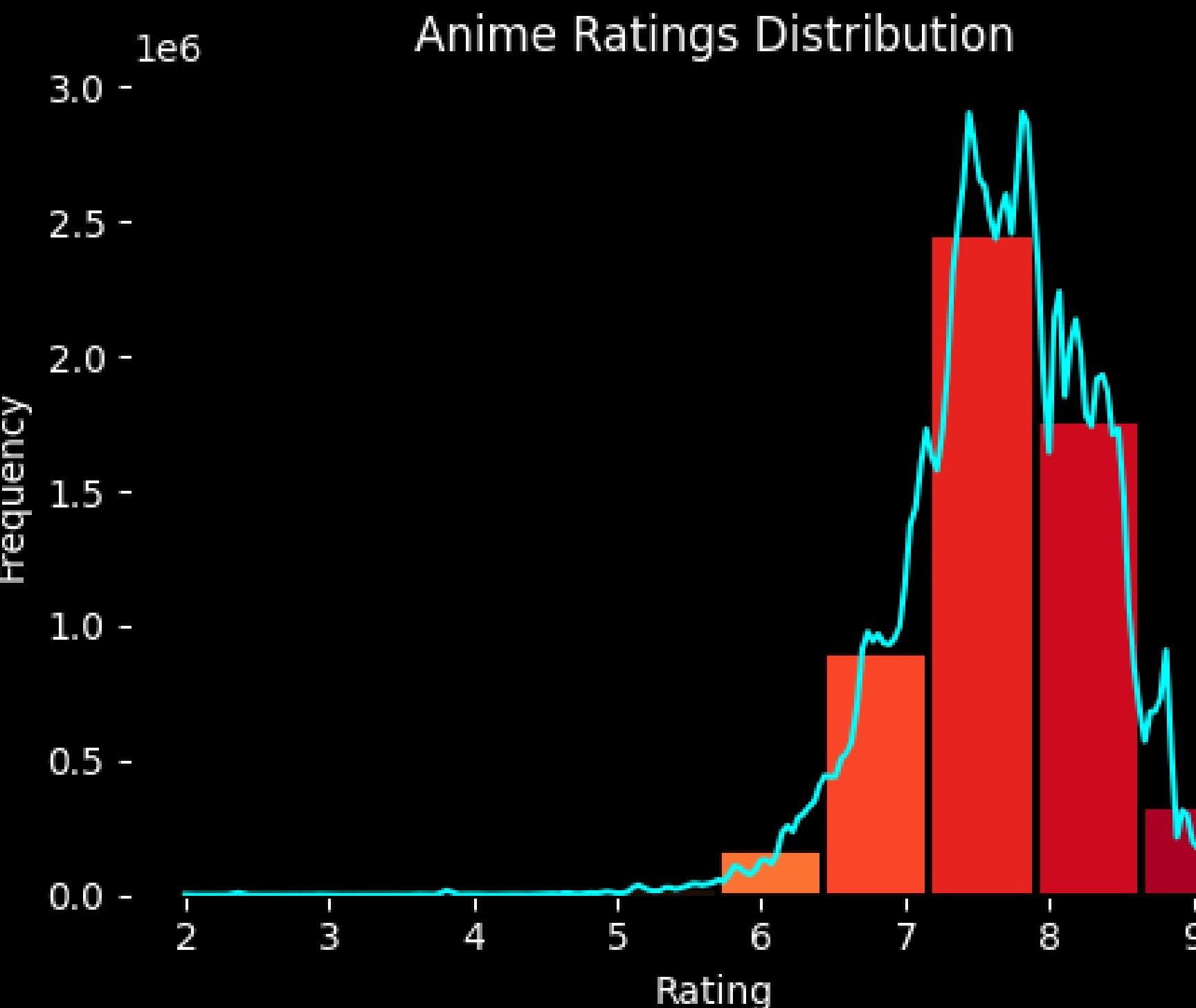
## Insights



- Users have varied opinions
- **Most ratings = 7 and 8**
- indicating generally positive feedback.

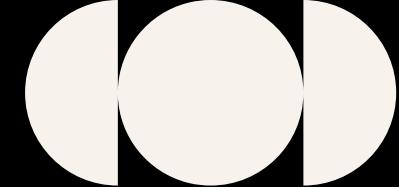


# Exploratory Data Analysis



## Insights

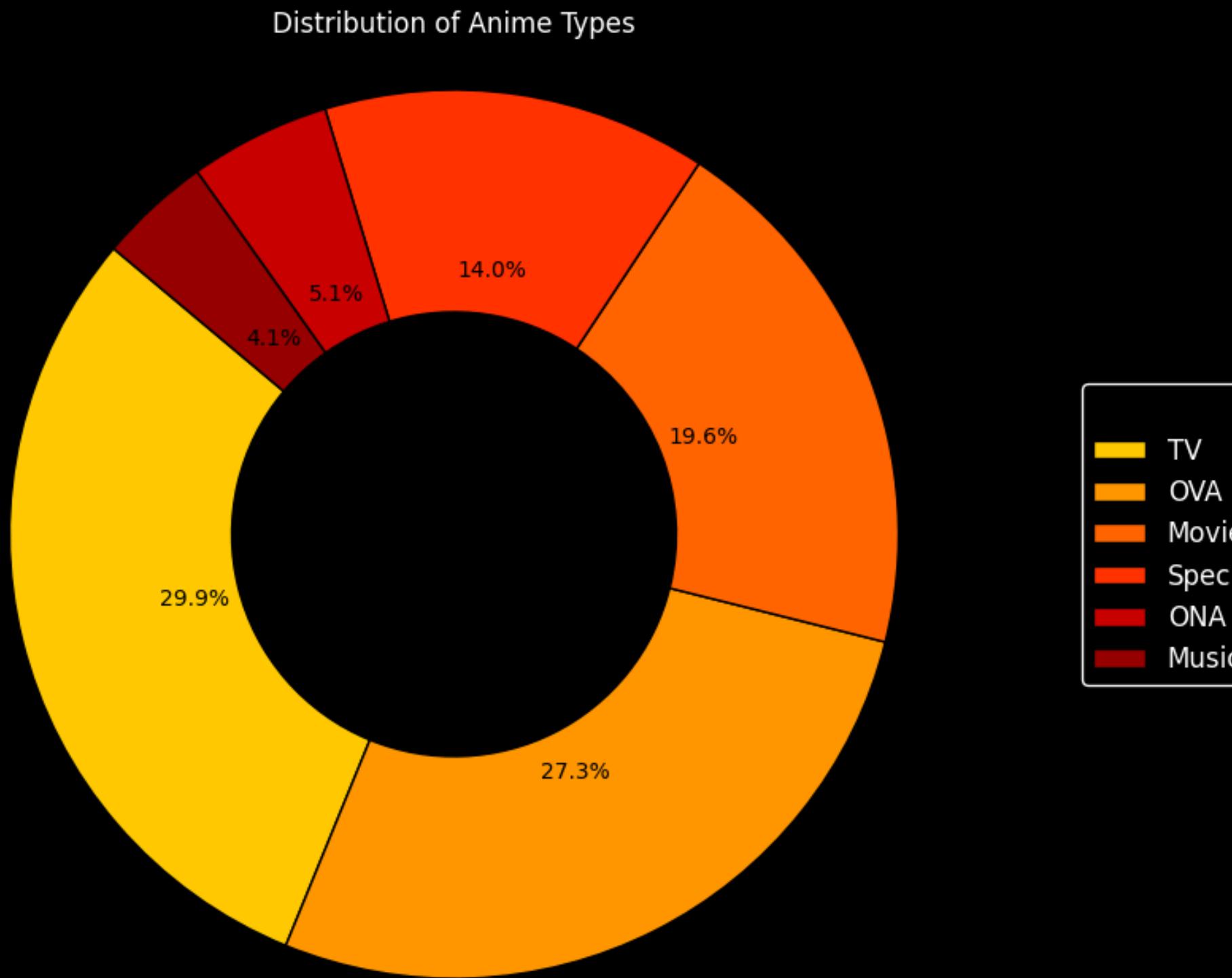
- Most anime ratings are mostly **higher** than **lower**.
- **Most ratings** = around 7
- Fewer very high ratings.



# Exploratory Data Analysis



## Insights

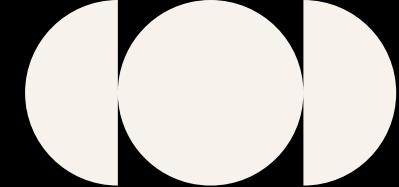


### TV Anime:

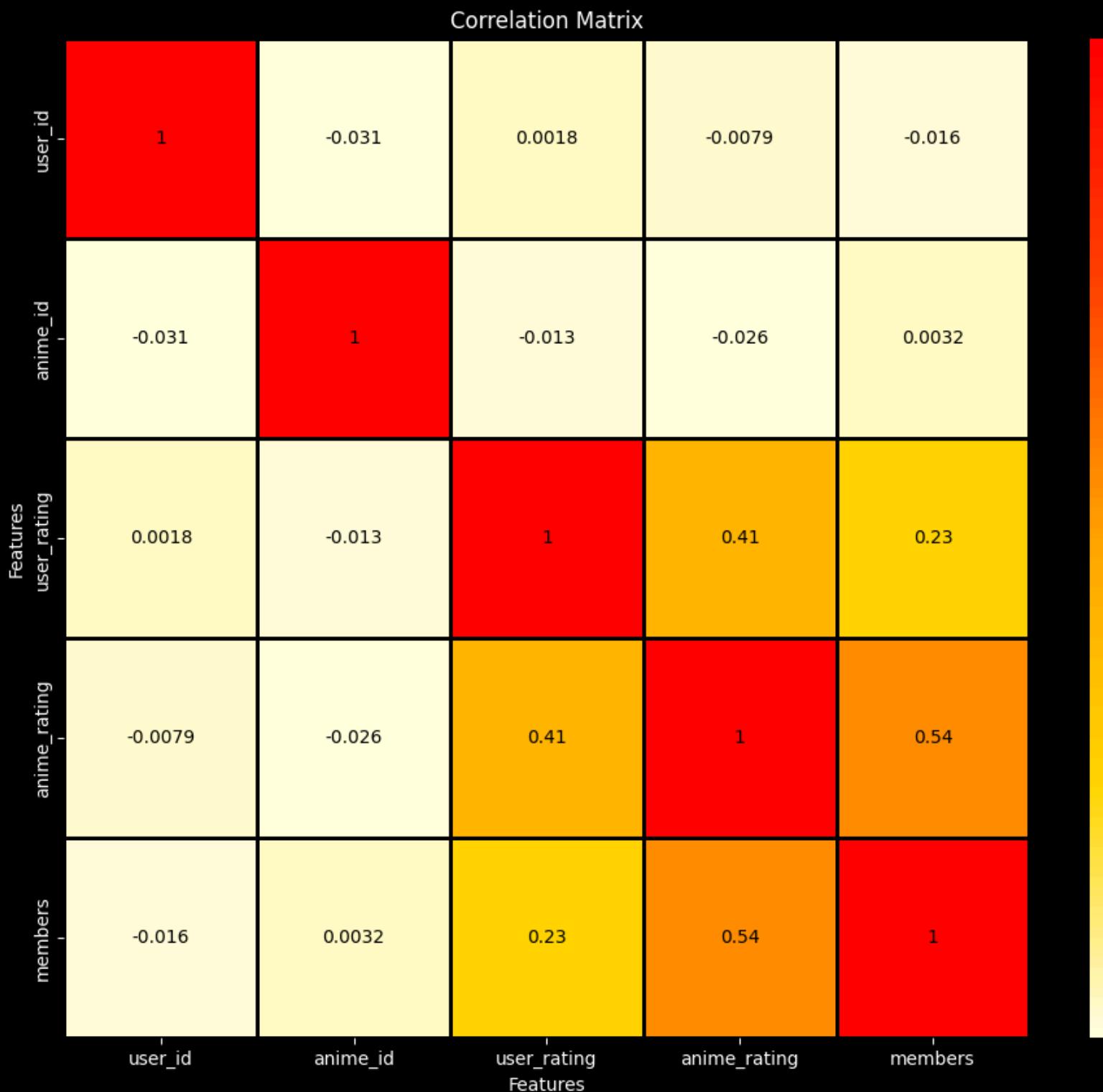
- most common type
- preference for episodic storytelling.

### Music:

- least common type
- due to niche appeal and rarity in the dataset.



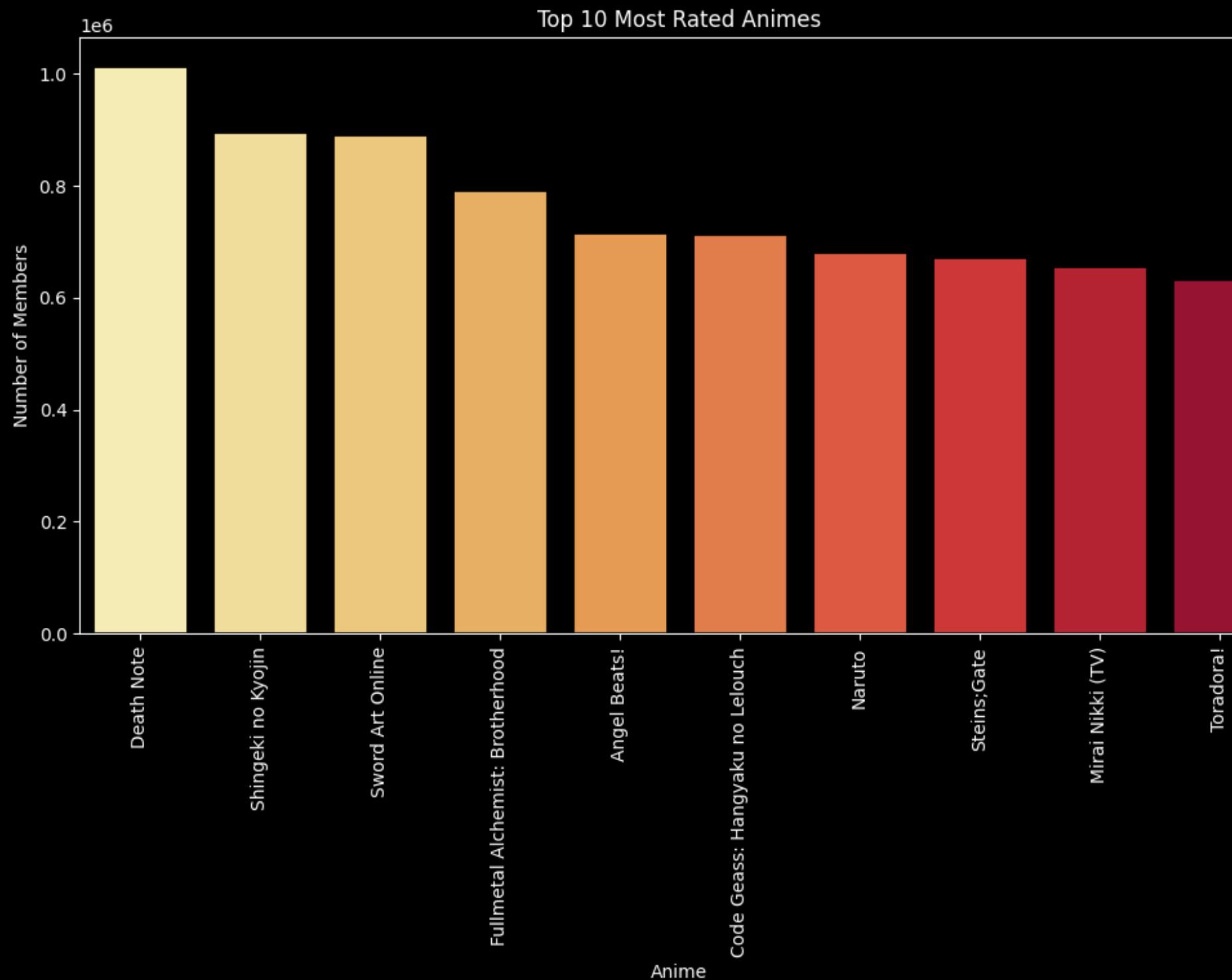
# Exploratory Data Analysis



## Insights

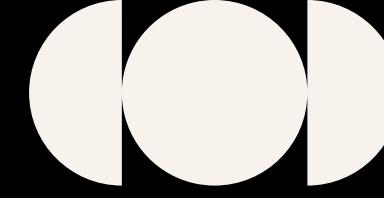
- Minimal Relationships
- Moderate Correlation:
  - anime ratings with user ratings & members
- Negligible Impact:
  - User IDs & anime IDs on ratings or popularity.

# Exploratory Data Analysis



## Insights

- **Top Anime:** Death Note has highest membership
- **Trend:** Membership decreases down to Toradora.



# Exploratory Data Analysis



# Insights

- Top Genres: Action, Comedy, and Adventure
  - Demographic Focus: Shounen and Shoujo show focus on young audiences.



# Preprocessing

## **Why Preprocessing Matters:**

- Formats data correctly for accurate recommendations.

## **What Was Done:**

- Treated user\_id and anime\_id as categories.
- Converted data to the required format.



# Models

1

## Surprise Library Models

- **SVD (Singular Value Decomposition):** A matrix factorization
- **Slope One: A simple collaborative filtering algorithm.**
- **NMF (Non-negative Matrix Factorization):** Another matrix factorization technique used for collaborative filtering.



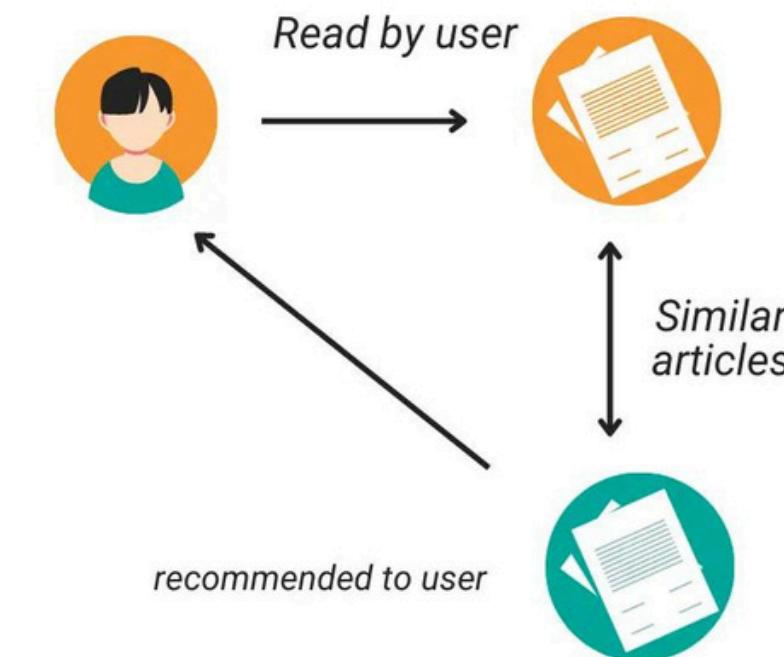
# Models

2

## Content Based Models

- **TF-IDF Vectorizer:** Used to convert textual data into numerical vectors
- **Cosine Similarity:** Used to measure similarity between two TF-IDF vectors

### CONTENT-BASED FILTERING

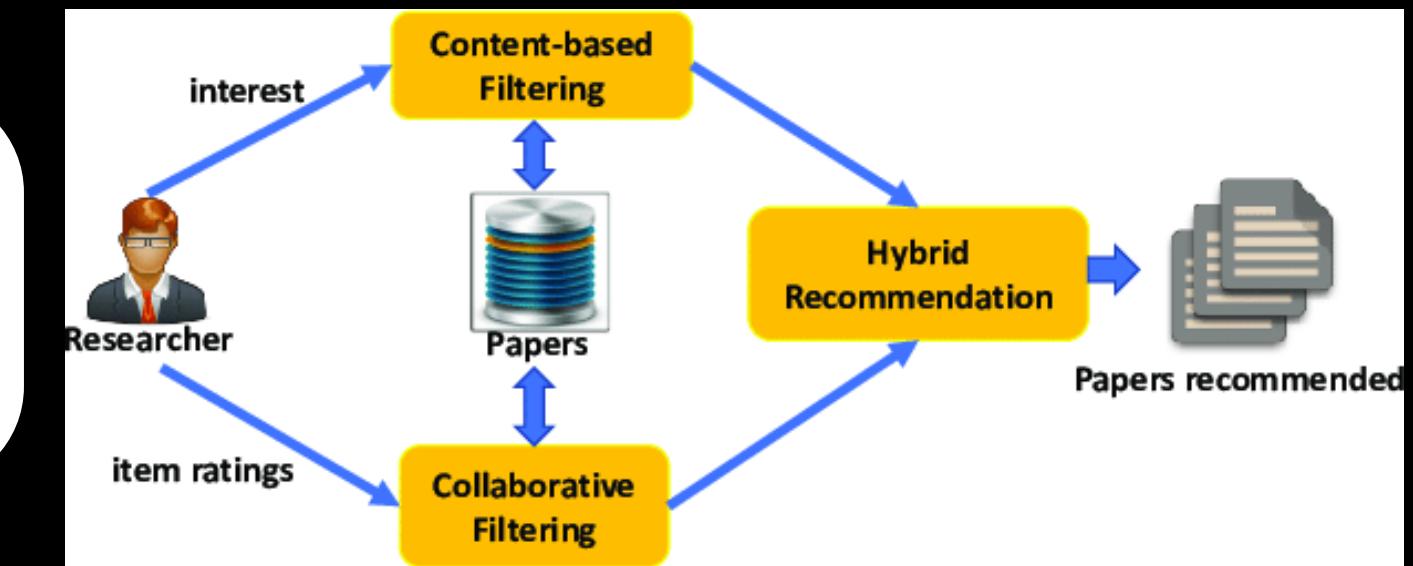


# Models

3

## Hybrid Recommendation System

A hybrid approach combining the predictions from NMF and the content-based model using cosine similarity.



## Hyper-parameter Tuning

**RandomSearchCV:**

- No significant improvement

# Best Model

## Metrics used

- Root Mean Square Error (RMSE)
- Assess the accuracy of top-N recommendations

## Selected Best Model

- The Hybrid Model (NMF + Content-Based) is the best performer
  - Diverse Recommendations
  - Personalization:
    - user's past behavior (collaborative filtering)
    - content similarity of items (content-based filtering)

# Streamlit

# Recommendations

01

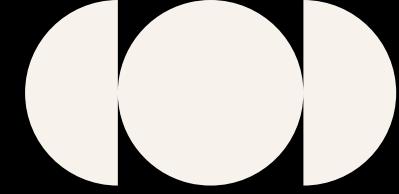
- Implement a Hybrid Recommendation System.
- Personalization and Diversity.

02

- Enhance Data Quality and Feature Engineering.
- User Feedback Integration

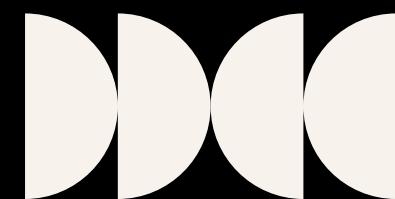
03

- Address the Cold-Start Problem.
- Optimize and Evaluate Models



# CONCLUSION

- The project has developed an effective system that captures user preferences accurately, using a hybrid approach with NMF to predict ratings and suggest anime titles.
- This system streamlines the content discovery process for anime fans, reducing the time spent searching for new series and increasing user engagement.





# THANK YOU

for your time and attention any questions ??

