GROUP 1

DELVIN OSORO ENID KIBET LOGAN NJIRU WINNIEFRED KINYUMU DENNIS LIMO

GAME OF THRONES

Play

My List

Watch Season 8 Now

The Night King and his army of the dead finally breach the walls of Winterfell, casting a pall of terror over the living. Just as all hope seems lost, a shocking betrayal throws the battle into chaos, leaving the fate of the realms hanging in the balance.

Group 1 Gs recommend to you:

GROUP 1

Project Overview

In this project, we built a movie recommendation system using the MovieLens dataset, applying collaborative and content-based filtering. We cleaned and analyzed the data, evaluated models, and found SVD to be the best for personalized recommendations.

Next Slide

OBJECTIVE:

- Analyzing movie ratings from the MovieLens dataset
- Using collaborative filtering and content-based filtering
- Recommending the top 5 movies for users

WHY IT MATTERS:

- Streaming platforms have vast movie libraries
- Users experience choice overload
- Personalized recommendations improve engagement

GROUP 1

BUSINESS UNDERSTANDING

In this project, we built a movie recommendation system using the MovieLens dataset to help users discover movies that match their preferences. We applied collaborative and content-based filtering, cleaned and analyzed the data, and evaluated multiple models. SVD proved the most effective, enabling personalized recommendations based on user ratings.

Dataset Overview

This project aimed to create a personalized movie recommendation system using the MovieLens dataset, which comprises 100,836 ratings, 3,683 user-generated tags, and 9,742 movies.

The system utilizes collaborative and content-based filtering techniques to analyze user ratings, movie genres, and tags to accurately predict and suggest movies tailored to individual preferences.

Users **No duplicates**

610

Ratings **Varies per user**

100,836

Tags **Describes Movie**

3,683

Movies **No Duplicates**

9,742

GROUP 1

Data CLEANING & PROCESSING

To ensure data quality, we checked for duplicates, missing values, and outliers, confirming that the dataset was complete and reliable. We retained low ratings as they provided valuable insights into user preferences. Key transformations included converting timestamps, extracting release years, and splitting multi-genre entries for detailed analysis. These steps ensured a clean, structured dataset for accurate movie recommendations.

01 Handling Duplicates

We checked for duplicate records across all datasets, including movies, ratings, and tags. No duplicate entries were found, ensuring that each record in the dataset was unique and that the integrity of the data was maintained.

02 Handling Missing Values

Missing values can negatively impact analysis and model performance. We verified that all key columns, such as movie titles, genres, and user ratings, had no missing values. This confirmed the completeness of our dataset and eliminated the need for imputation.

03 Managing Outliers

While some ratings appeared as outliers, particularly very low ratings, we decided to retain them. These outliers provided valuable insights into user preferences, helping to capture strong negative opinions, which are essential for improving recommendation accuracy.

GROUP 1

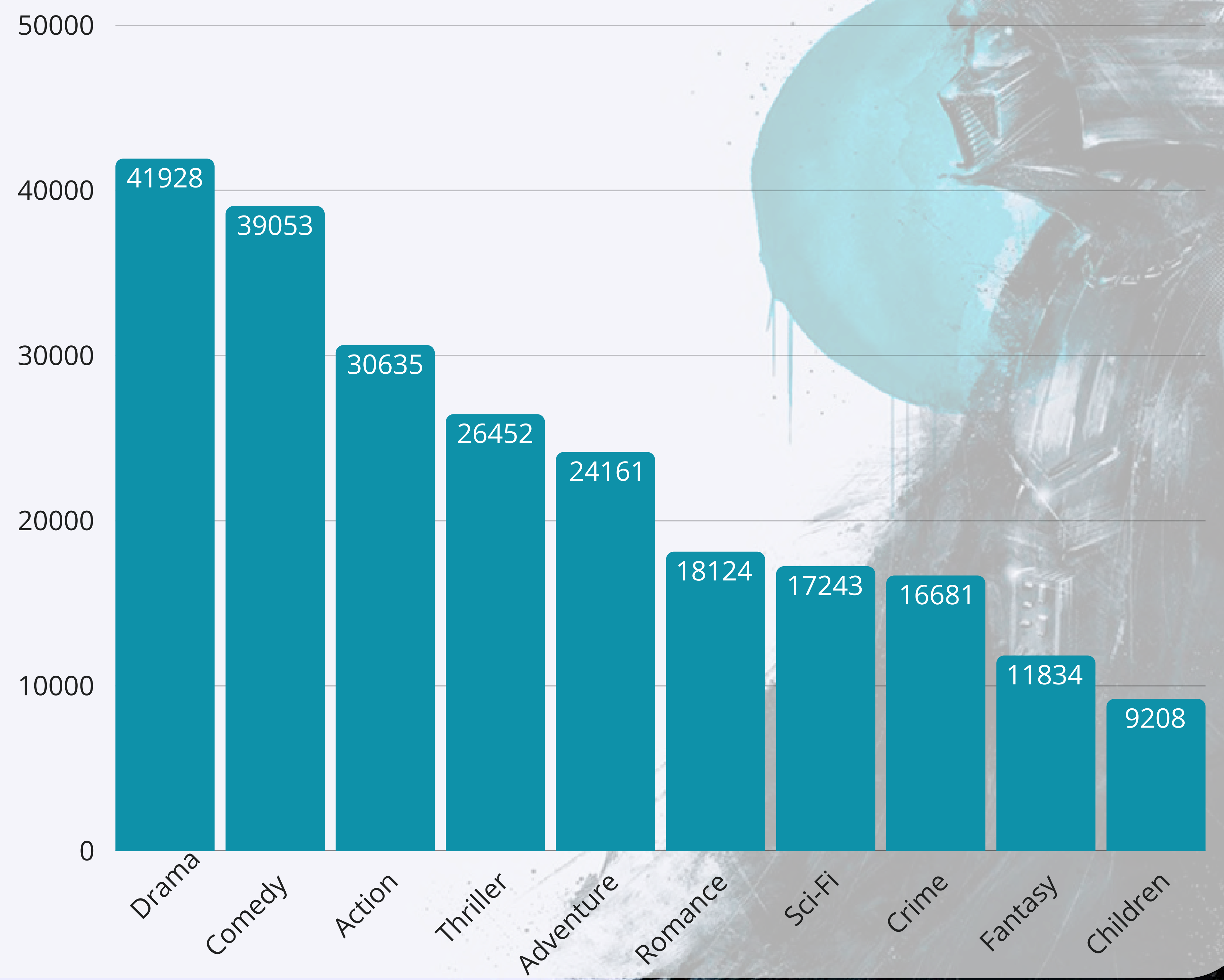
EXPLORATORY DATA ANALYSIS

Most Popular Genres

About Data

Most Popular Genres: Drama, Comedy, and Action received the highest number of ratings, showing they are widely watched and rated. This suggests a strong audience preference for engaging storytelling, humor, and action-packed entertainment.

Learn More

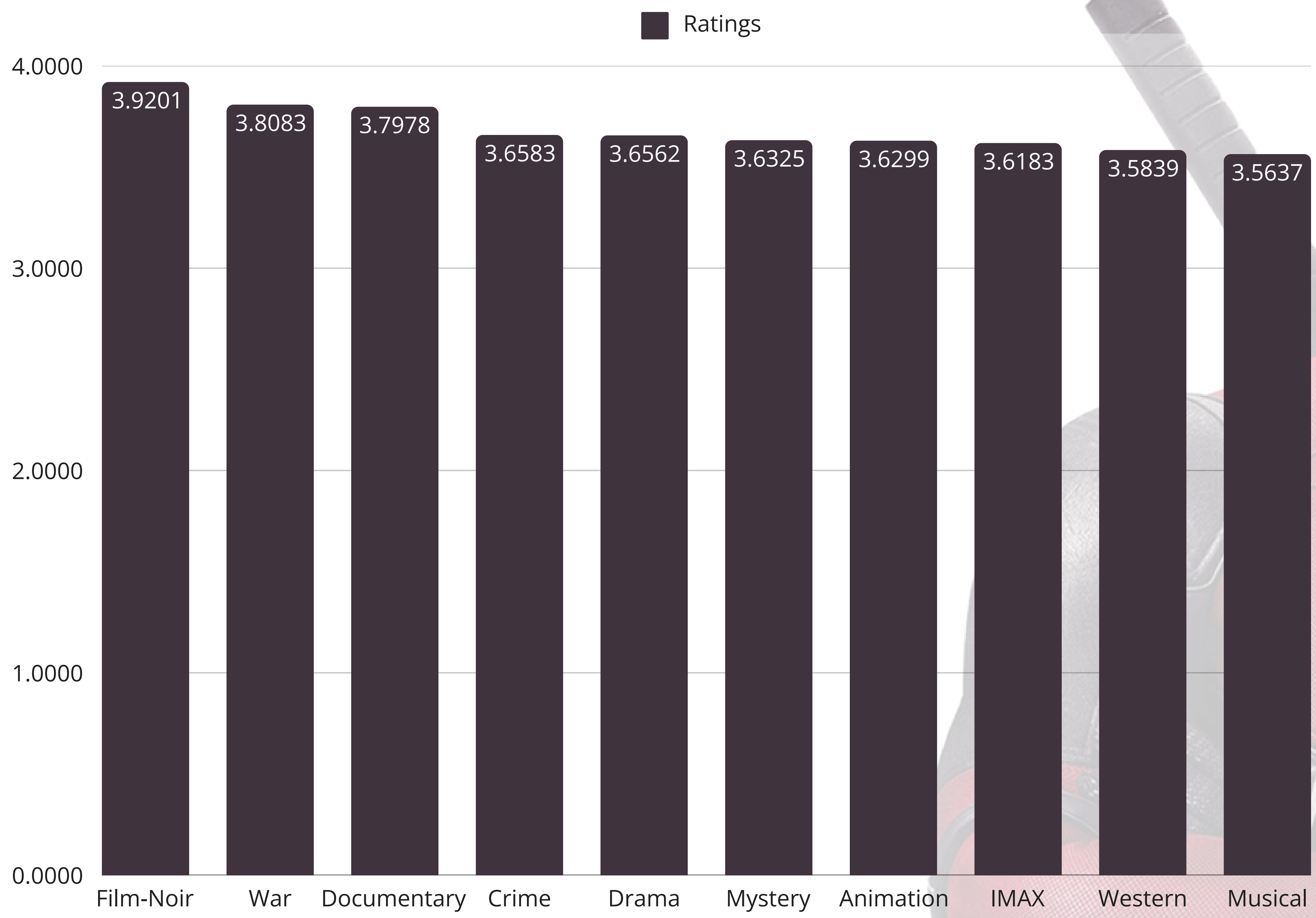


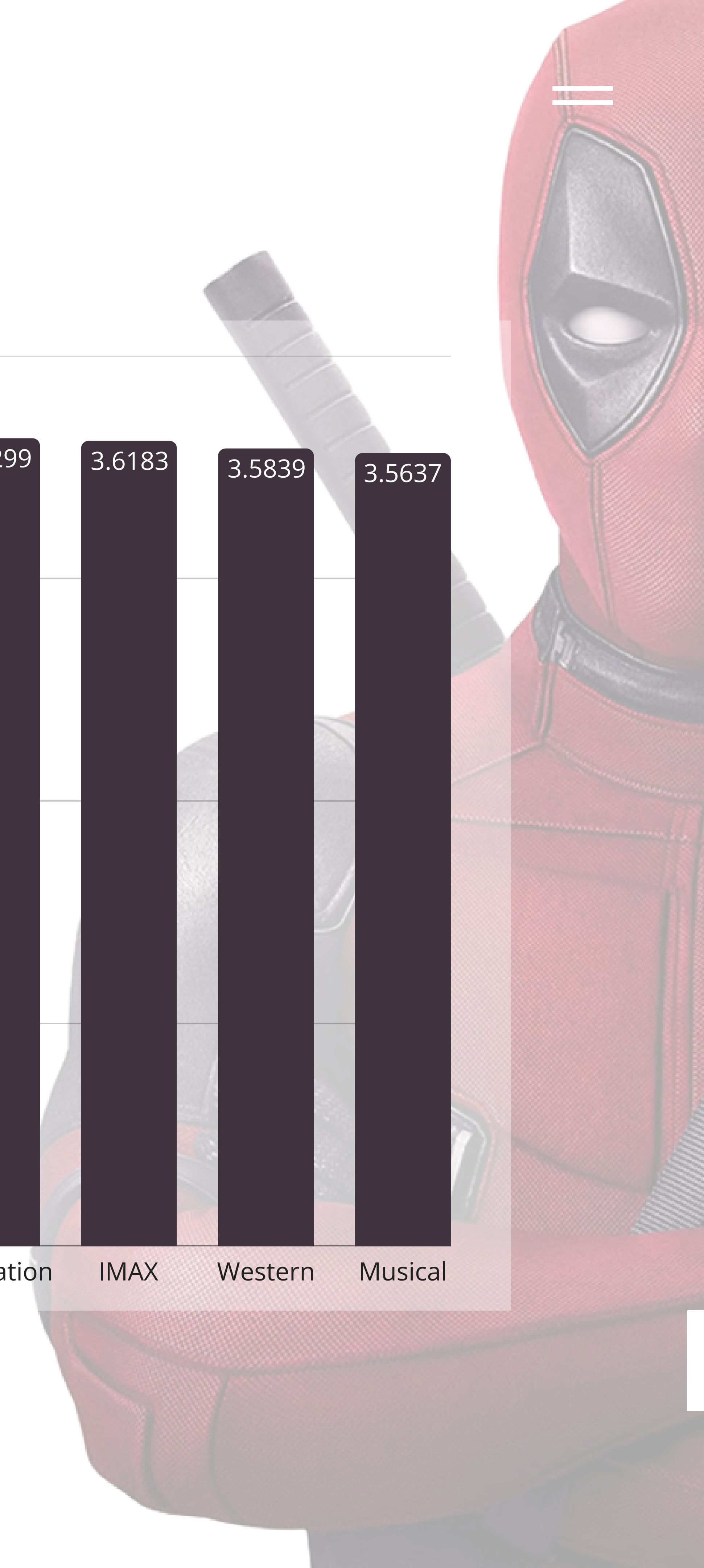
GROUP 1

Highest Rated Genres

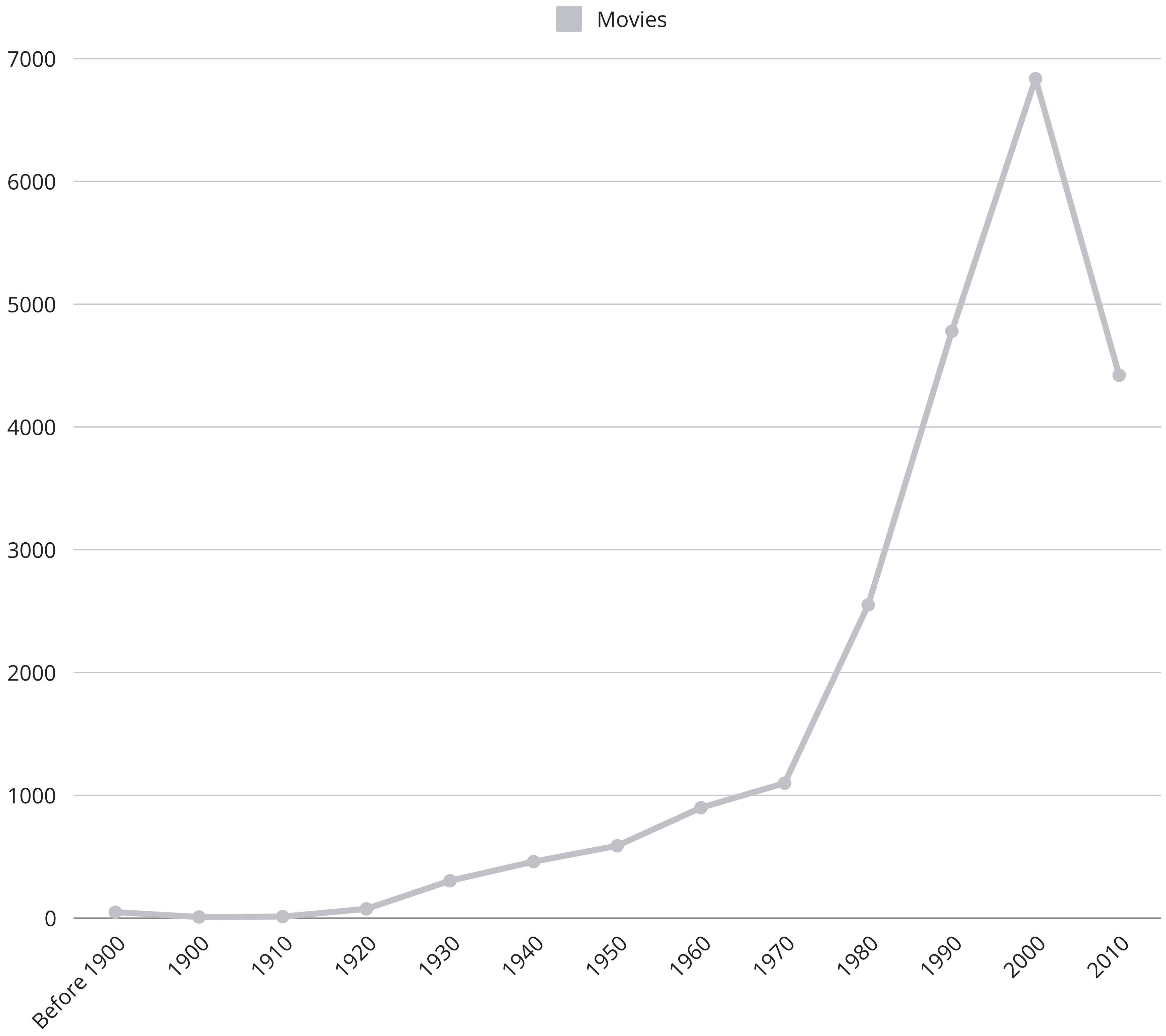
Film-Noir, War, and Documentary had the highest average ratings, indicating they are well-received despite being less mainstream. Their strong ratings suggest that niche audiences highly appreciate their storytelling depth and realism.

Learn More





GROUP 1



Movie Production Trends:

Movie production saw a steady rise after 1970, driven by technological advancements and increasing demand. The fluctuation in production in recent decades either reflects the impact of digital streaming, market shifts, and evolving audience preferences or indicates the data was obtained early in the 2010s.

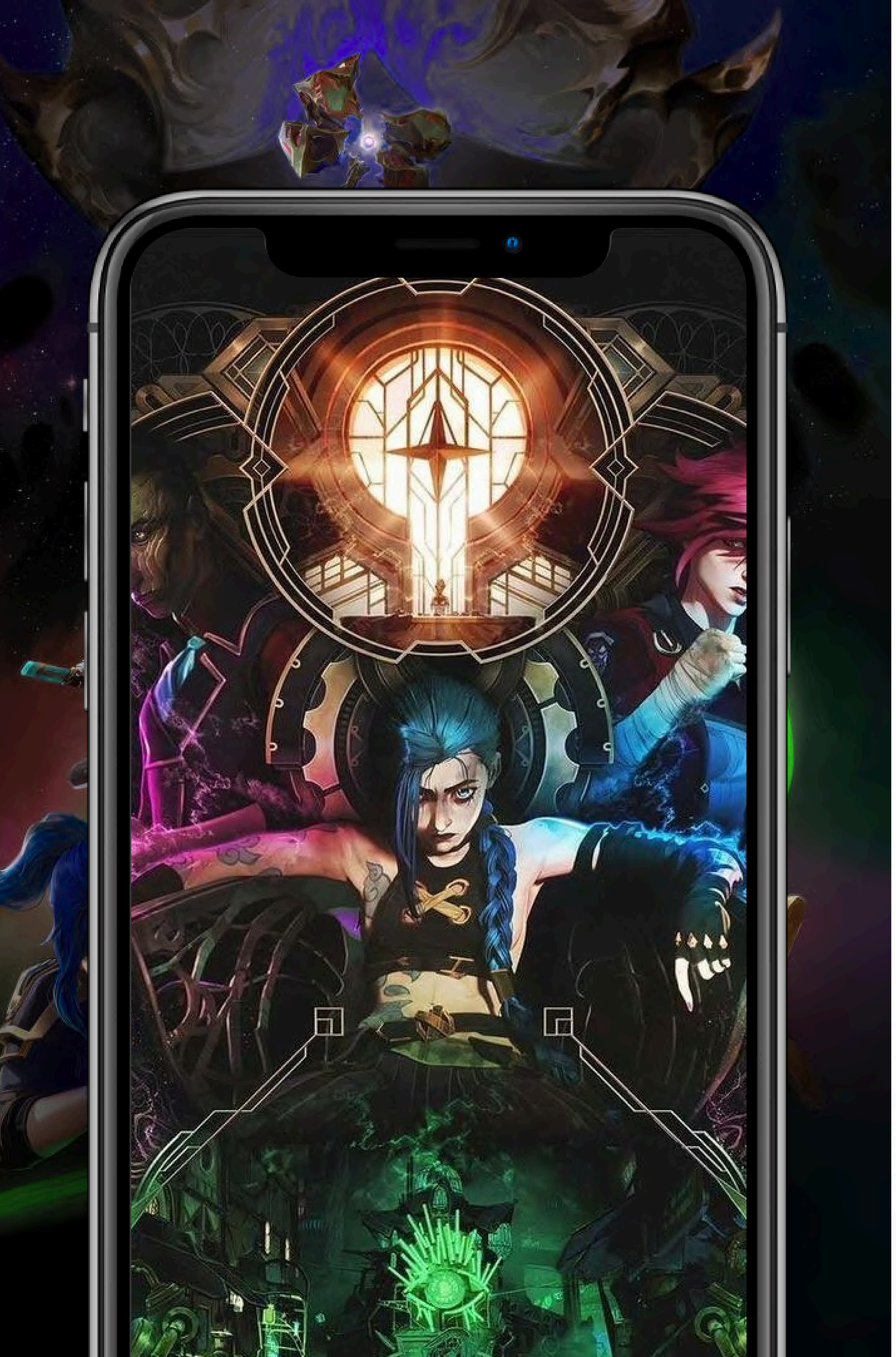
RECOMMENDATION
Model Selection

One of the key challenges in building the recommendation system was data sparsity, as most users rated only a small subset of movies. To address this, we used collaborative filtering, which predicts missing ratings based on patterns in user behavior.

We evaluated multiple models to determine the best-performing approach:

- SVD (Singular Value Decomposition): Achieved the lowest RMSE (0.85), making it the most accurate model.
- KNNBasic & KNNBaseline: Had higher error rates, making them less suitable for precise recommendations.

SVD emerged as the optimal model due to its ability to capture hidden patterns in user-movie interactions, leading to better predictions and personalized recommendations.



WHERE BRANDS START

The recommendation system works by first gathering user input, where users rate a few movies from their preferred genre. This helps the model understand their preferences. Using SVD, the system then predicts the top 5 movies that the user is most likely to enjoy based on their ratings and similar users' preferences. For example, if a user rates 5 movies, the system analyzes these ratings and provides 5 personalized recommendations tailored to their tastes, ensuring a more engaging and relevant viewing experience.The cold start problem occurs when a recommendation system struggles to provide accurate suggestions for new users with no prior ratings or new movies with little to no feedback.

Addressing the Cold Start Problem:

- New Users: The system guides new users to rate a few movies to gather initial data for personalized recommendations.
- New Movies: Content-based filtering (genre, tags, similarities to popular films) is used to recommend new movies with limited user data.

EXAMPLE *Recommendation*

```
      movieId                                    title  genres  \
13393    40578  Sword of the Beast (Kedamono no ken) (1965)  Action

      Release_year  Decade
13393         1965    1960
You rated this movie: n
Not Rated
      movieId          title  genres  Release_year  Decade
20336   134246  Survivor (2015)  Action          2015    2010
You rated this movie: 4
      movieId                                       title  genres  \
18104    95519  Dragon Ball Z: Bojack Unbound (Doragon bôru Z ...  Action

      Release_year  Decade
18104         1993    1990
You rated this movie: 3
      movieId                              title  genres  Release_year  \
15956    69654  Prison Break: The Final Break (2009)  Action          2009

      Decade
15956    2000
You rated this movie: 3
     movieId                     title  genres  Release_year  Decade
4980    2989  For Your Eyes Only (1981)  Action          1981    1980
You rated this movie: 4
```

```python
#Importing the function for displaying the recommended movies based on the user ratings using collaborative filtering
from Functions import recommend_movies

# Print recommended movies
recommendations = recommend_movies(user_ratings, movie_df, rating_df, svd, num_recommendations=5)

print(f"Recommended Movies based on collaborative filtering:\n{recommendations}")
```

```
Based on the 4 movies you've rated, here are some recommendations tailored to your preferences. These movies have high pre
cted ratings, suggesting you might enjoy them!
Recommended Movies based on collaborative filtering:
       movieId                                              title  \
10354     7121                                 Adam's Rib (1949)
21572   170705                            Band of Brothers (2001)
10151     7008  Last Tango in Paris (Ultimo tango a Parigi) (1...
12289    27156  Neon Genesis Evangelion: The End of Evangelion...
5319      3224           Woman in the Dunes (Suna no onna) (1964)

       predicted_rating
10354          4.648134
21572          4.584489
10151          4.580171
12289          4.572542
5319           4.540651
```

Interpretation

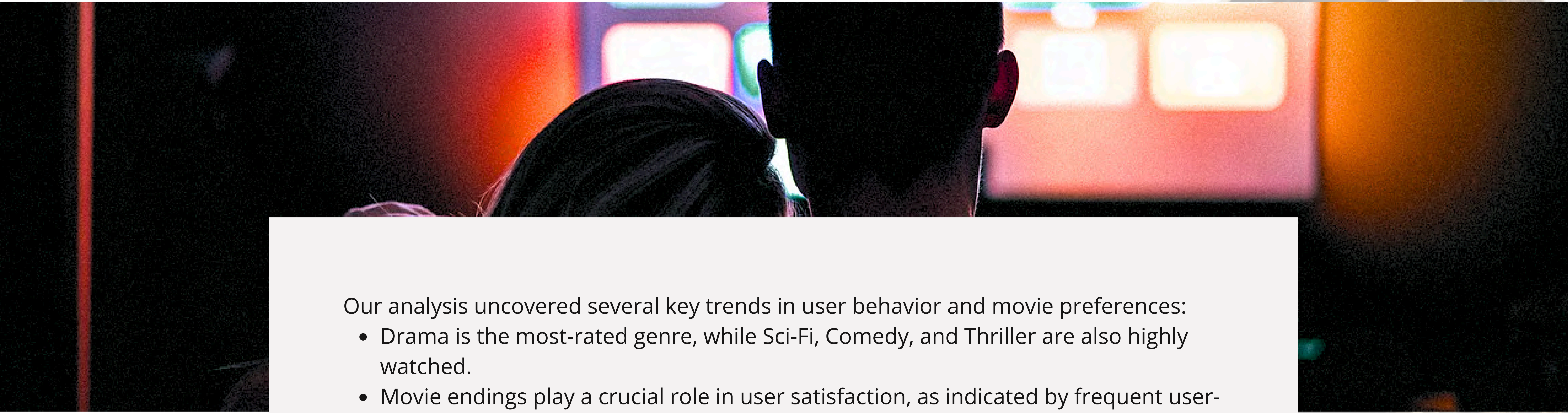
The above code displays the recommended ratings to the user based on the different movies they rated.

Addressing Cold-start Problem

A user rates five movies from their preferred genre, providing the system with initial input to understand their preferences. Using SVD, the model analyzes these ratings and recommends five personalized movies based on similar users' preferences and viewing patterns.

GROUP 1

KEY FINDINGS & INSIGHTS



Our analysis uncovered several key trends in user behavior and movie preferences:

- Drama is the most-rated genre, while Sci-Fi, Comedy, and Thriller are also highly watched.
- Movie endings play a crucial role in user satisfaction, as indicated by frequent user-generated tags.
- Personalized recommendations significantly enhance engagement by reducing choice overload and guiding users toward relevant content.

Key Findings & Insights

Our analysis uncovered several key trends in user behavior and movie preferences:

- Drama is the most-rated genre, while Sci-Fi, Comedy, and Thriller are also highly watched.
- Movie endings play a crucial role in user satisfaction, as indicated by frequent user-generated tags.
- Personalized recommendations significantly enhance engagement by reducing choice overload and guiding users toward relevant content.

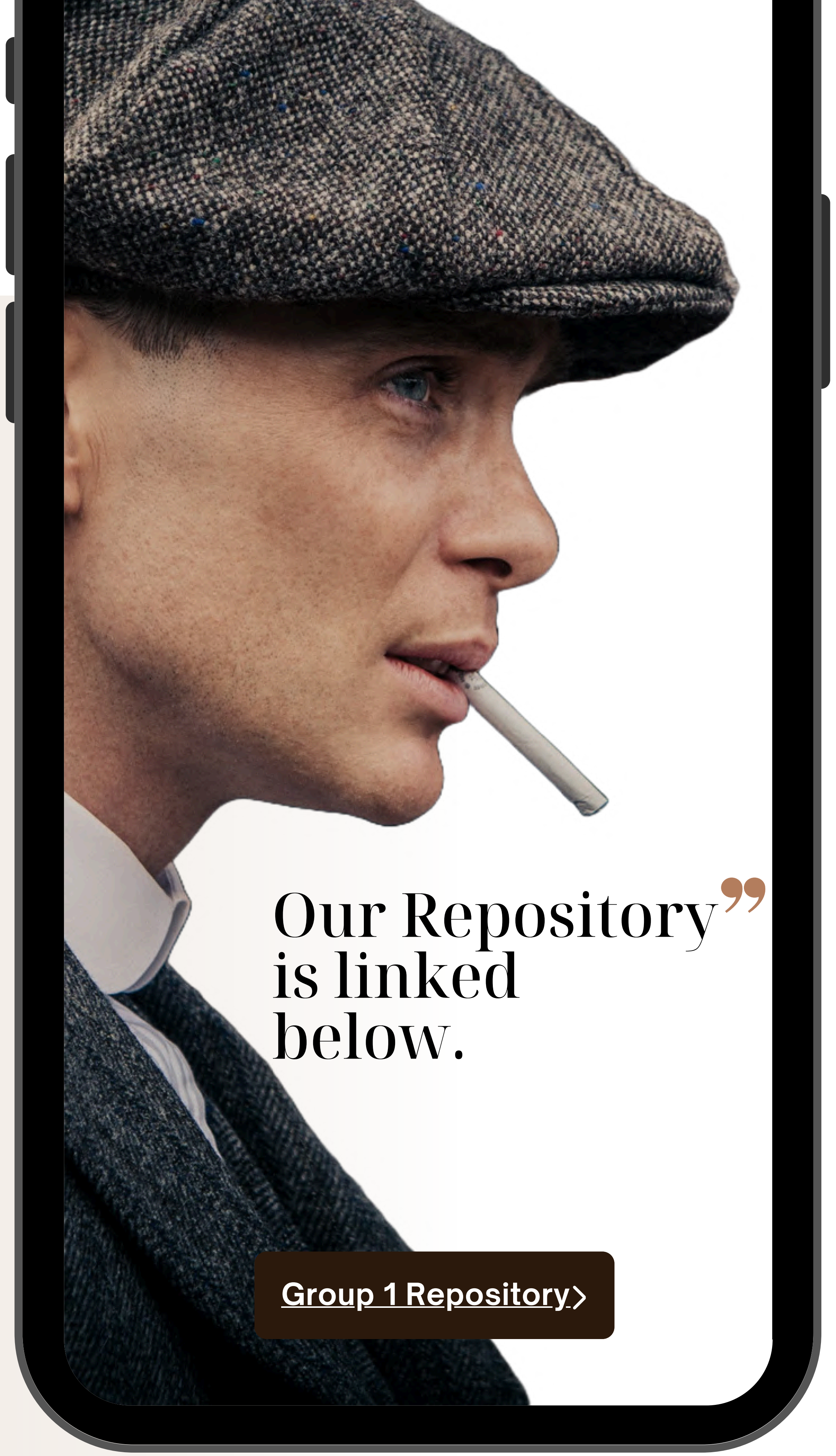
Conclusion & future Work

CONCLUSION

- The recommendation system effectively personalizes movie suggestions based on user preferences.
- Collaborative filtering with SVD proved to be the most accurate and reliable approach.

FUTURE WORK:

- Implement a hybrid recommendation system combining collaborative and content-based filtering for improved accuracy.
- Explore deep learning models to enhance user preference predictions and recommendation quality.



Thank You