Movie Recommender Model

PHASE 4: GROUP 12

MEMBERS;

1. DANIEL WAHOME.
2. PURITY GITONGA.
3. BRIAN KARIITHI.
4. CAROLINE GESAKA.
5. JAMES NJOROGE.

Group 12

INTRODUCTION

In the era of digital content consumption, personalized recommendation systems play a crucial role in enhancing user experience and engagement. Leveraging the MovieLens dataset from the GroupLens research lab, we aim to build an effective movie recommendation system. This system will utilize collaborative filtering techniques to suggest movies based on user ratings, thereby providing personalized movie recommendations.

PROBLEM STATEMENT

Develop a recommendation system that provides top 5 movie recommendations to users based on their past ratings. The system should address the challenge of sparsity in user ratings and the cold start problem for new users.

- Purpose of the Study
  - Goal: Develop a robust and accurate movie recommendation system, evaluate and optimize it.
  - Importance: Enhance user experience and engagement by providing personalized movie recommendations.
  - Objectives: Improve recommendation accuracy, understand user behaviour, and explore model scalability.

- Scope of the Analysis
  - Dataset: MovieLens dataset (small)
    - url: https://grouplens.org/datasets/movielens/latest/
  - Columns: ‘movieId', 'imdbId', 'tmdbId', 'userId_x', 'rating', 'timestamp_x', 'title', 'genres', 'userId_y', 'tag', 'timestamp_y’

2. DATA PREPARATION AND CLEANING

- Data Cleaning
  - Handling Missing Values and Dropping Irrelevant
- Data Transformation
  - Feature Engineering: Created new features such as rating frequency, average user rating, and genre popularity.
- Normalization:
  - Scaled ratings between 0 and 1 using Min-Max scaling to standardize input for model training.

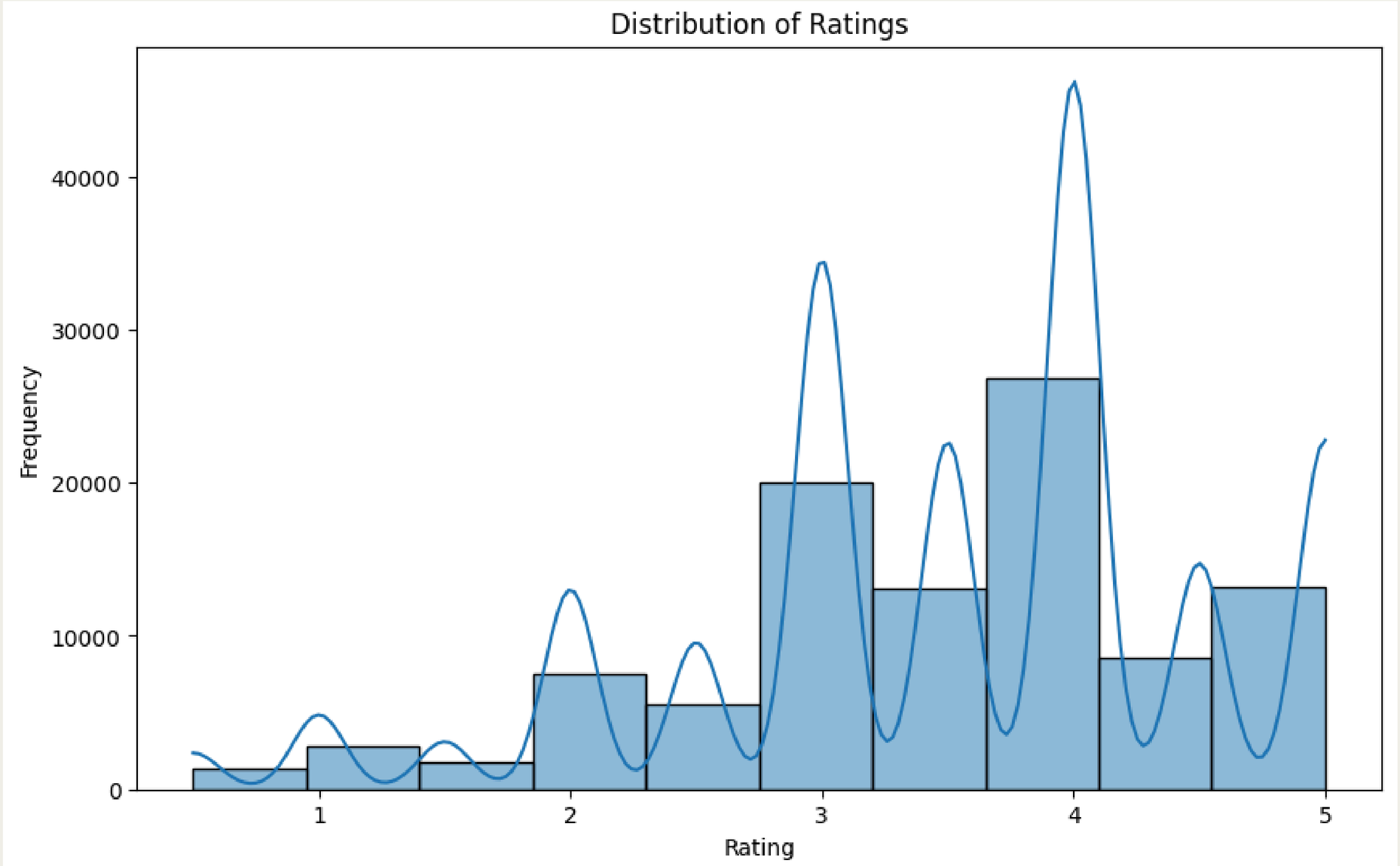
EXPLORATORY DATA ANALYSIS (EDA)

- Summary Statistics:
  - Mean rating: 3.84
  - Standard deviation: 1.02
- User and Movie Statistics:
  - 'Pulp Fiction (1994)' is the most rated film of genre; Comedy|Crime|Drama|Thriller and tagged as sci-fi

	title	genres	tag
count	285783	285783	285783
unique	9737	951	1589
top	Pulp Fiction (1994)	Comedy|Crime|Drama|Thriller	sci-fi
freq	55567	56864	55076

VISUALIZATION

Distribution of Ratings

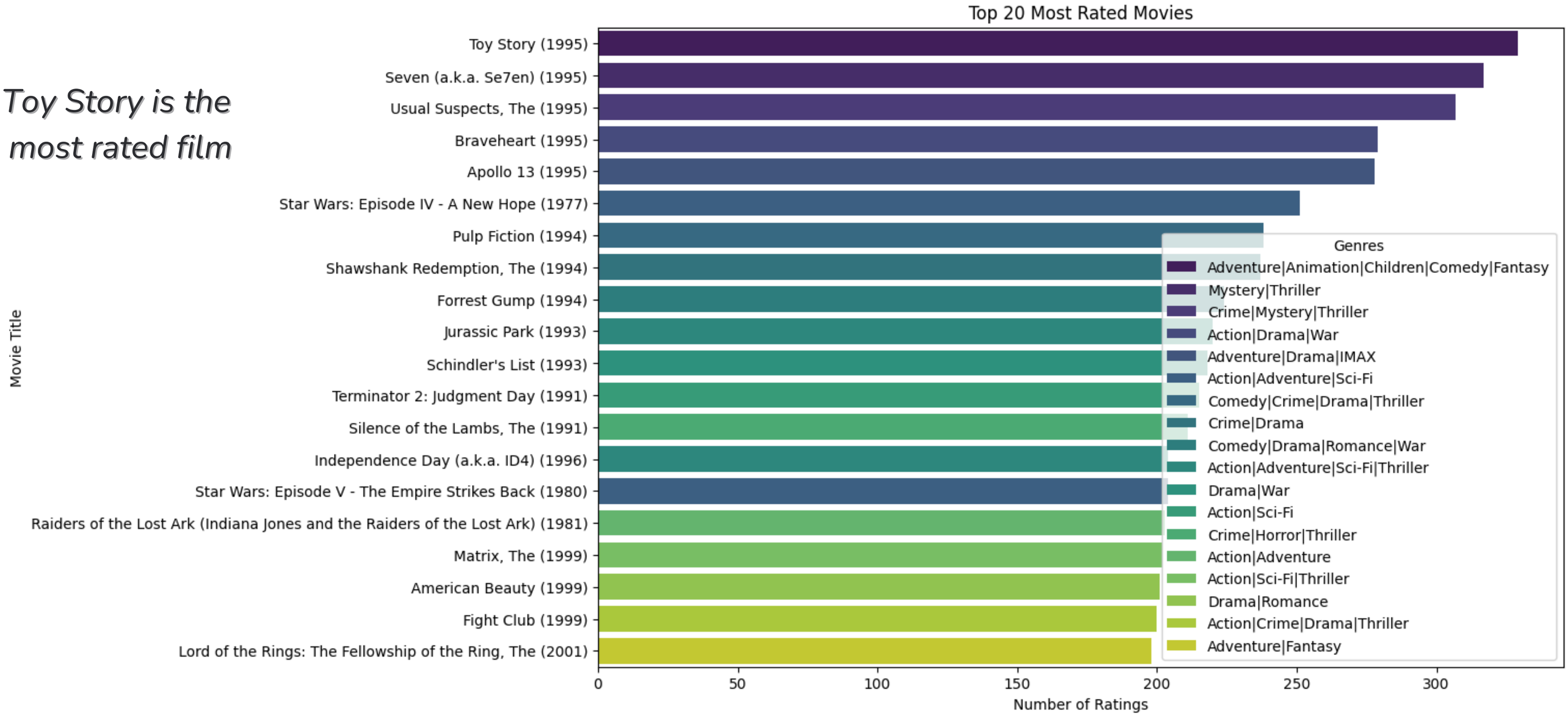


Most films are rated
at a 4
followed by a 3.5

VISUALIZATION

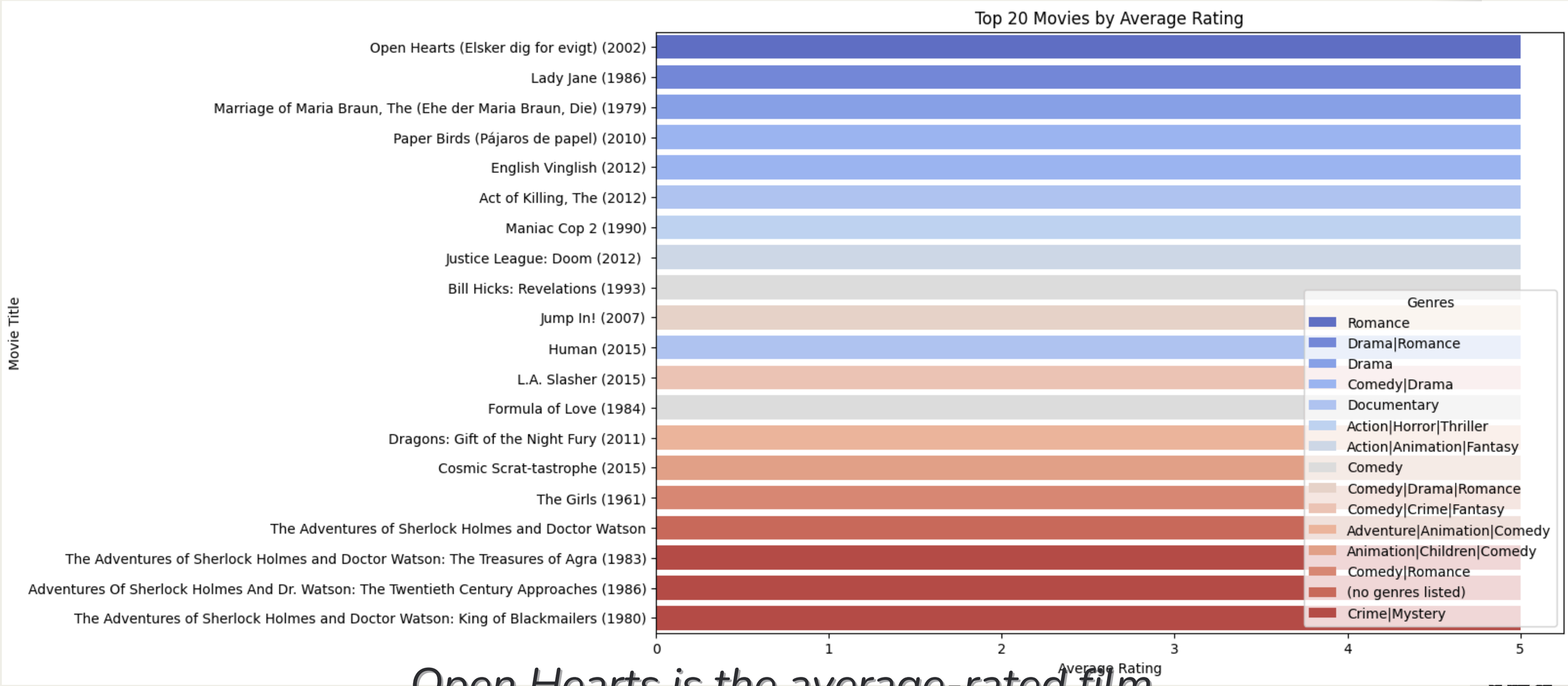
Ratings per Movie

Toy Story is the most rated film



VISUALIZATION

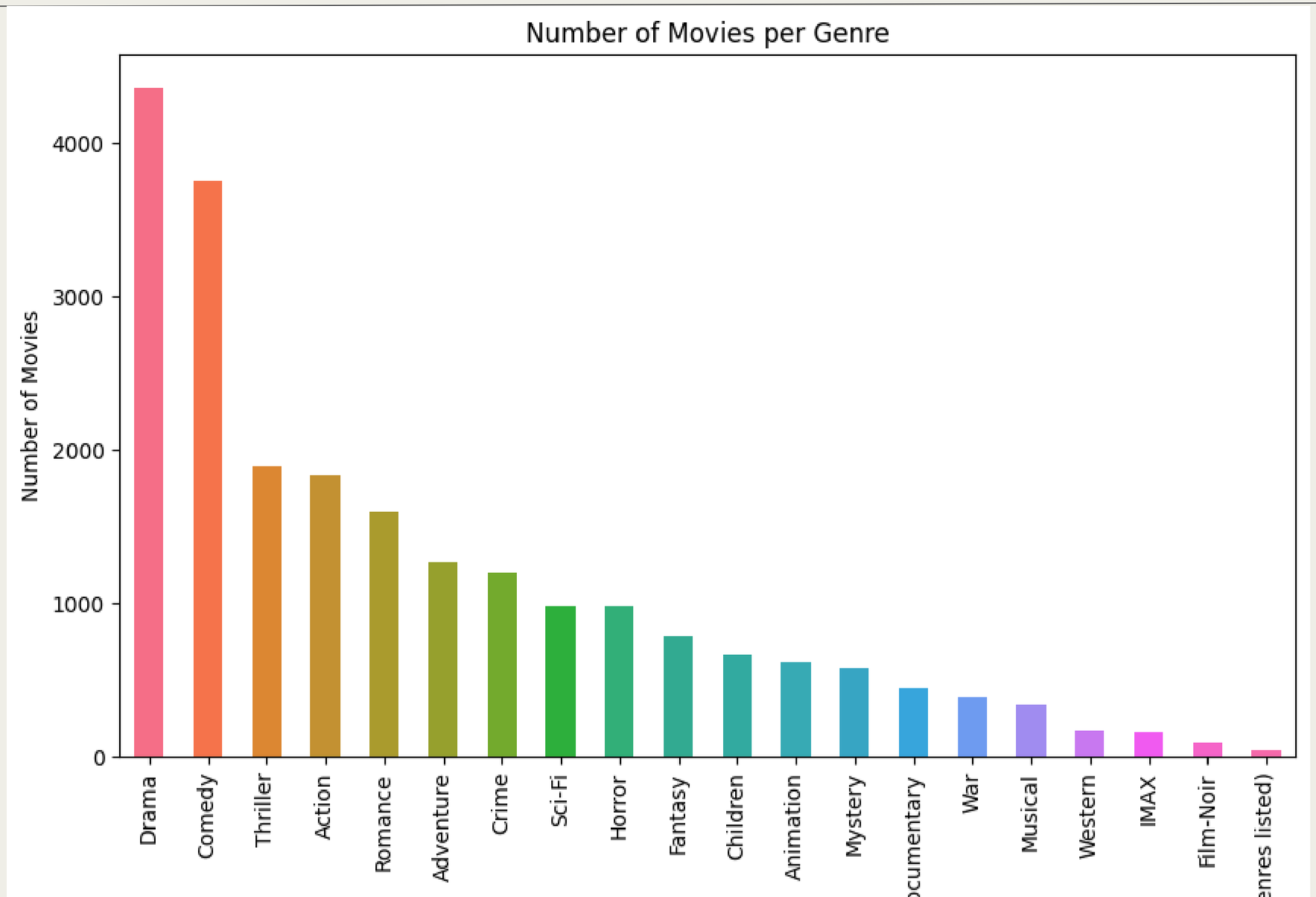
Ave-Ratings per Movie



Open Hearts is the average-rated film

VISUALIZATION

No. Movies per Genre

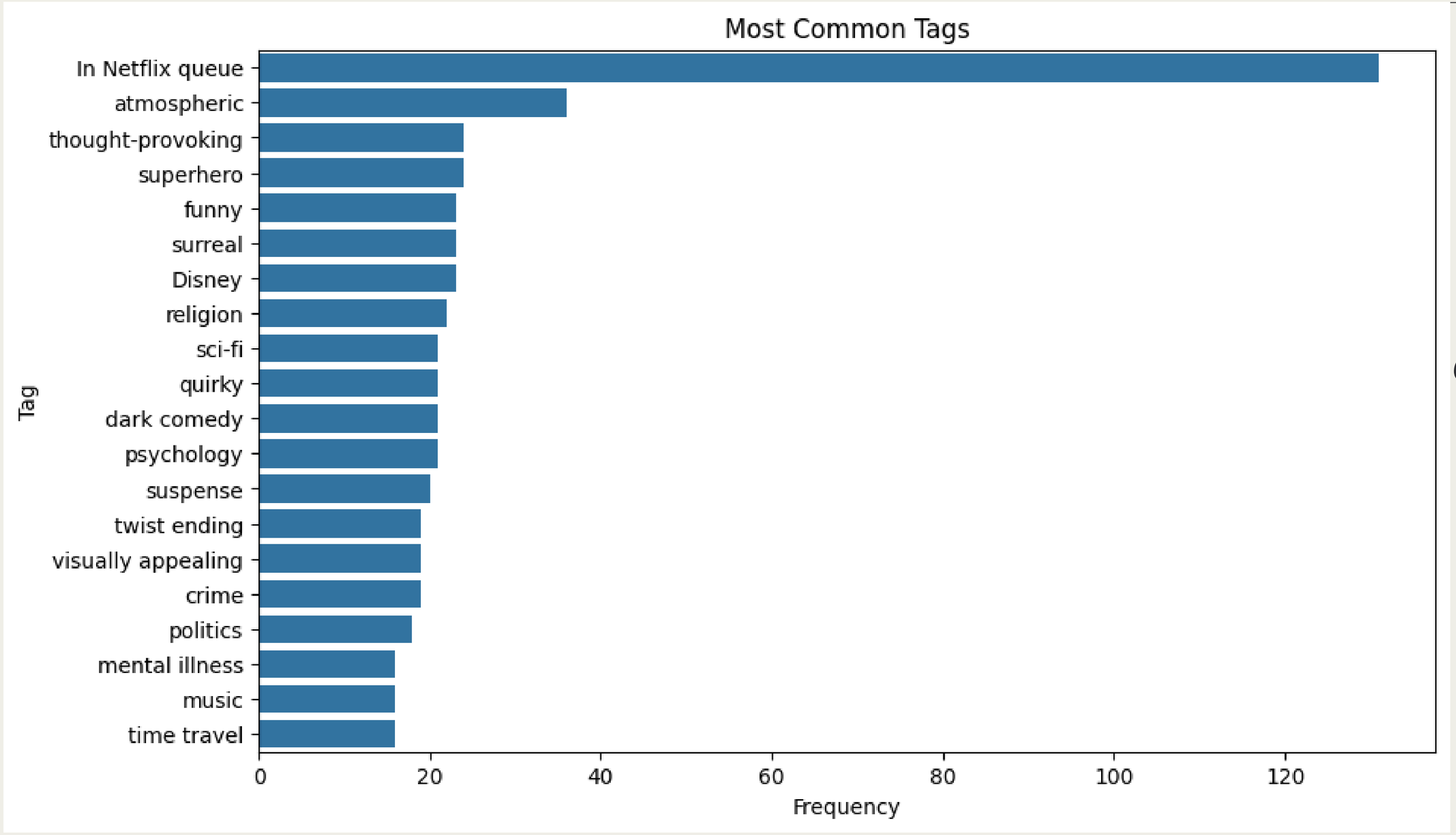


Most Movies are Drama.

Group 12

V I S U A L I Z A T I O N

Most common Tags



Most tags are in Netflix queue which shows most users utilise Netflix as a streaming platform

Group 12

MODELLING

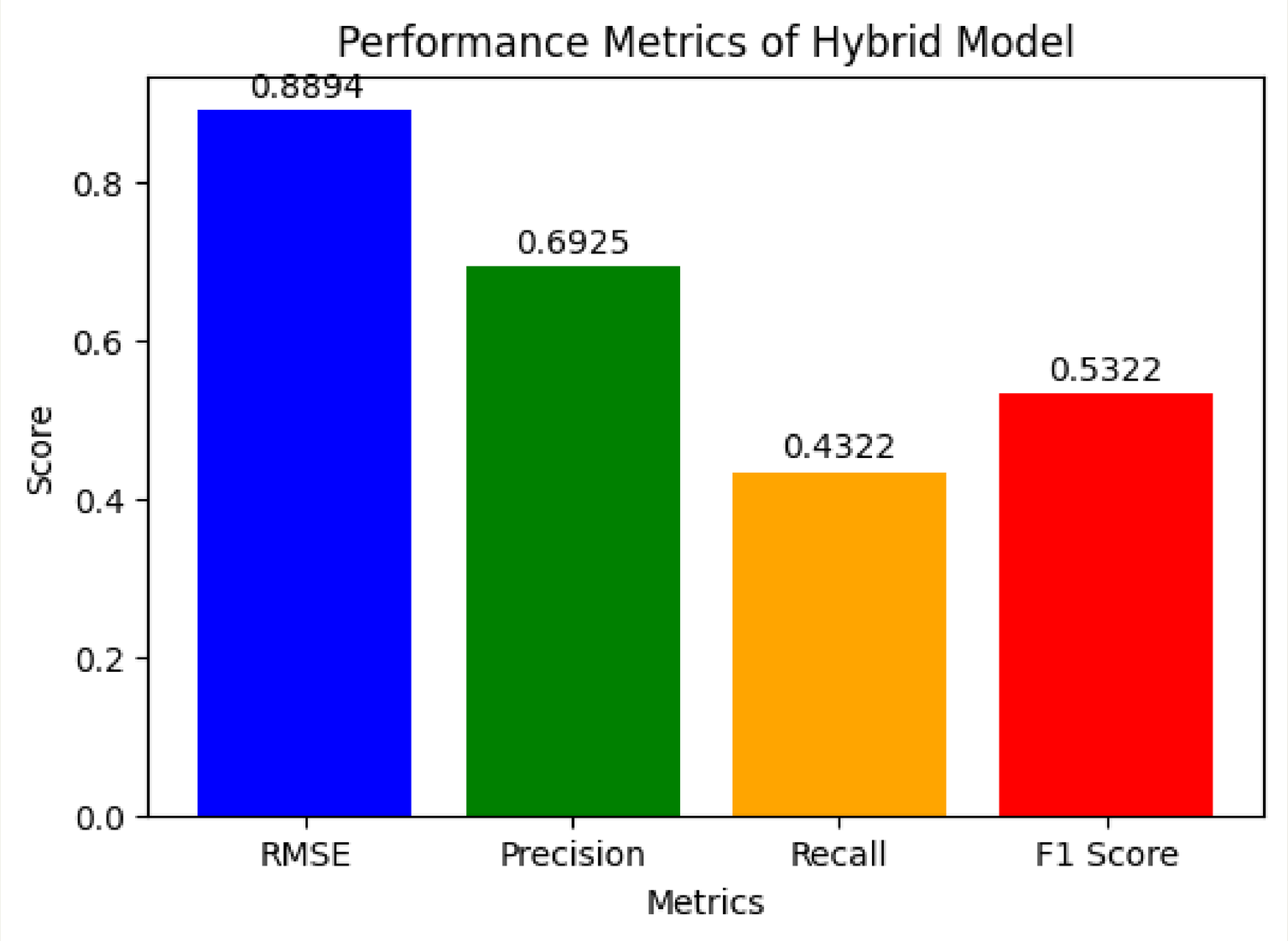
Algorithms Used:

- Baseline Model: Recommend the most popular movies and calculate baseline RMSE.
- User-Based Collaborative Filtering: Using KNN for user-based recommendations.
- Item-Based Collaborative Filtering: Using KNN for item-based recommendations.
- Matrix Factorization: Using SVD for recommendations.
- Hybrid Model: Combining user-based and item-based predictions.

MODELLING

- Data Splitting:
  - Train-test split with 80% training data and 20% test data.
- Cross-Validation:
  - Applied k-fold cross-validation to evaluate model performance.
- Performance Metrics
  - Evaluated using RMSE, MAE, and precision-recall metrics.

MODEL VIZ OF PERFORMANCE METRICS OF HYBRID MODEL



MODELLING EVALUATION

1. Most Popular Movies (Baseline)

RMSE: 1.0425

on average, the predictions are off by around 1.0425 units (which typically corresponds to the rating scale used.

2. User-Based Collaborative Filtering:

RMSE: 0.9562 **Precision: 0.6895** **Recall: 0.4325** **F1 Score: 0.5316**

predictions are more accurate than the baseline, reducing the average error in predictions.

successfully recommends relevant movies to users with a good balance between precision and recall.

MODELLING EVALUATION

3. Item-Based Collaborative Filtering:

RMSE: 0.9162 **Precision: 0.6574** **Recall: 0.4176** **F1 Score: 0.5108**
suggests further improvement in prediction accuracy compared to user-based CF and the baseline.
performs well but needs improving

4. SVD (Matrix Factorization):

RMSE: 0.8828 **Precision: 0.6911** **Recall: 0.4282** **F1 Score: 0.5288**
better accuracy than both CF methods and the baseline, indicating it can make more precise predictions.

MODELLING EVALUATION

3. Item-Based Collaborative Filtering:

RMSE: 0.9162 **Precision: 0.6574** **Recall: 0.4176** **F1 Score: 0.5108**
suggests further improvement in prediction accuracy compared to user-based CF and the baseline.
performs well but needs improving

4. SVD (Matrix Factorization):

RMSE: 0.8828 **Precision: 0.6911** **Recall: 0.4282** **F1 Score: 0.5288**
better accuracy than both CF methods and the baseline, indicating it can make more precise predictions. competitive performance with high precision and balanced recall, indicating effective recommendations.

MODELLING EVALUATION

5.Hybrid Model (Combination of CF and Content-Based):

RMSE: 0.8894 **Precision: 0.6925** **Recall: 0.4322** **F1 Score: 0.5322**

strong performance, slightly below SVD but still significantly better than baseline and CF methods.
performs similarly to SVD, demonstrating robustness in recommending relevant movies with high precision and recall.

RECOMMENDATIONS

Recommendations

- Model Improvements
  - Future Work: Incorporate additional features like user demographics and social interactions.
  - Enhance Algorithms: Experiment with deep learning models for better accuracy.
- Business Implications
  - Improved User Engagement: Tailored recommendations can increase user retention.
  - Strategic Recommendations: Focus on popular genres during certain times of the year.

C O N C L U S I O N

By leveraging the MovieLens dataset and implementing collaborative filtering techniques, this project aims to deliver a robust movie recommendation system. The insights gained from this system not only benefit users by providing personalized movie suggestions but also provide valuable learning in the field of recommendation systems.

GROUP 12 GITHUB REPO LINK

url:https://github.com/Daniel-Wahome/GROUP-12-PHASE-4.git

STREAMLIT RECOMMENDER DEPLOYMENT LINK

url:https://movielense.streamlit.app/

Thank you!

FROM GROUP 12 MEMBERS