MOVIE **RECOMMENDATION SYSTEM**

by:
Eric Mutua
Amani Wanene
Sharon Mungai
Winny Chepkoech

CONTENT

INTRODUCTION

This analysis aims to provide personalized movie recommendations to users on a streaming platform based on their movie preferences and rating history. In the highly competitive streaming industry, offering relevant and tailored content is essential to increasing user engagement and reducing churn. By leveraging the MovieLens dataset from the GroupLens research lab at the University of Minnesota, this project seeks to create a recommendation system that enhances the user experience and drives long-term engagement.

BUSINESS CONTEXT

A streaming service is looking to improve user satisfaction and engagement by offering personalized movie recommendations. Despite having a large library of content, many users are not engaging with the platform as expected, resulting in lower average watch times and higher churn rates. The company wants to develop a movie recommendation system that can provide users with a tailored list of movies based on their past ratings and viewing history, increasing the likelihood of engagement and retention.

DATA

The MovieLens Dataset, sourced from the GroupsLens research Lab at the University of Minnesota contains a colection of movie ratings and associated metadata. For this project we'll use the smaller dataset which contains 100,000 ratings and 3,600 tag applications applied to 9,000 movies by 600 users.

DATA OVERVIEW

Key Features:

- User ID: Represents unique users.
- Movie ID: Unique identifiers for movies.
- Rating: User ratings of movies (on a scale from 1 to 5).
- Tags: Keywords or phrases describing movie themes (e.g., “Action,” “Adventure”).
- Timestamp: Records when users rated the movie.

PROCESS STEPS

- DATA CLEANING
- EXPLORATORY DATA ANALYSIS
- MODELING

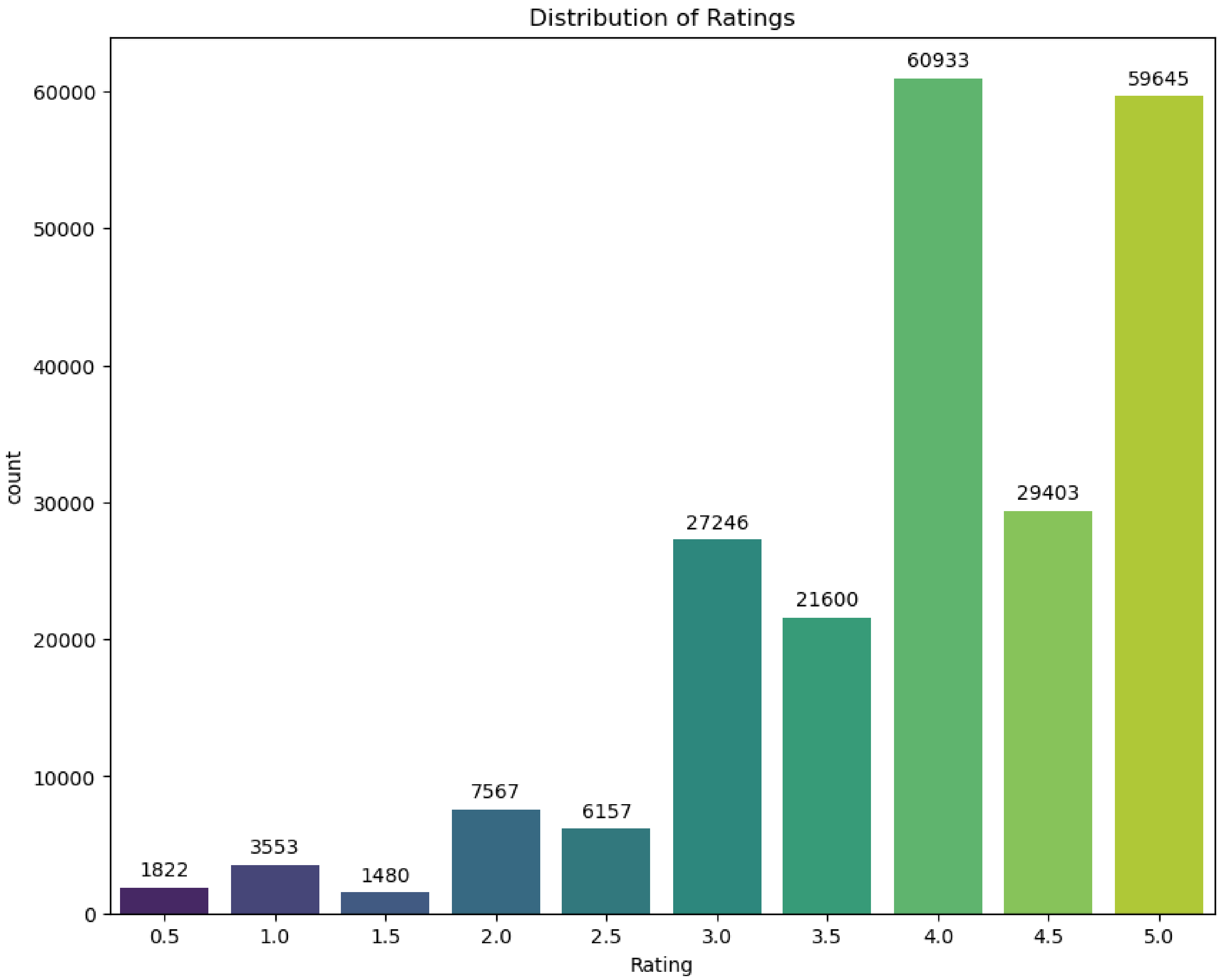
DATA CLEANING

DATA PREPARATION

- Checking for missing values.
- Checking for duplicates.
- Drop irrelevant columns.
- Feature engineering

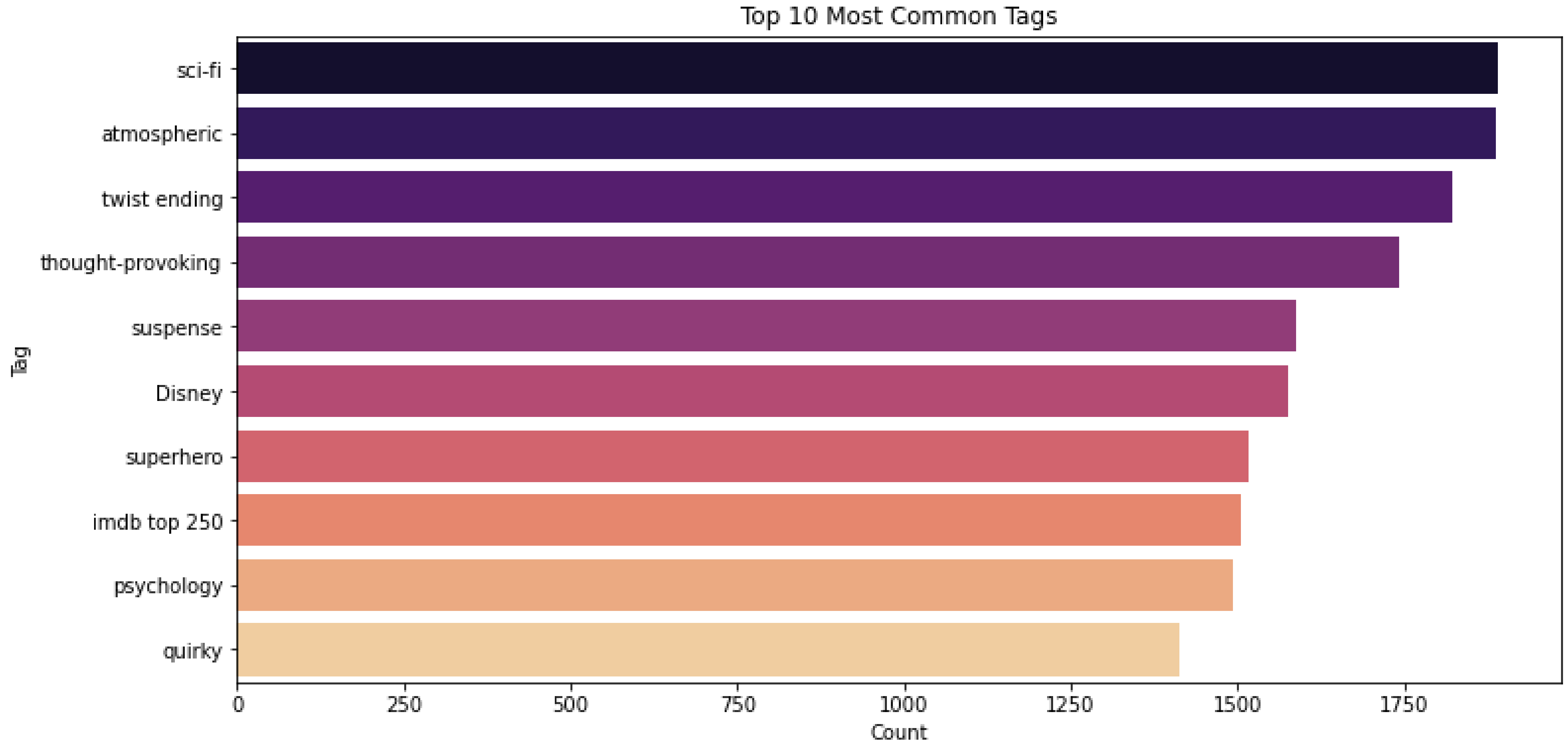
EXPLORATORY DATA ANALYSIS

DISTRIBUTION OF MOVIE RATINGS



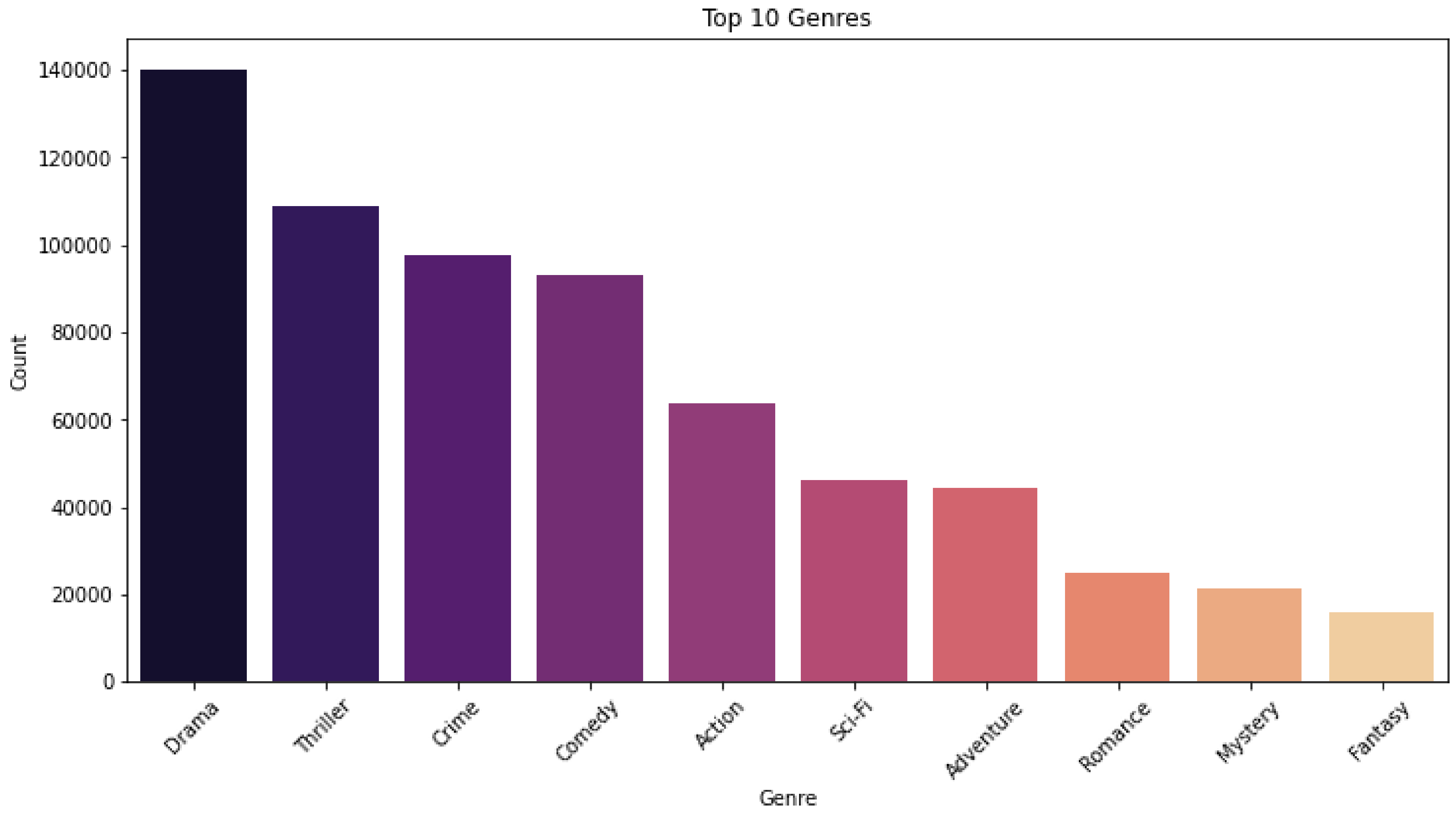
4.0(60933) and **5.0**(59645) are the most frequent ratings

DISTRIBUTION OF MOVIE TAGS



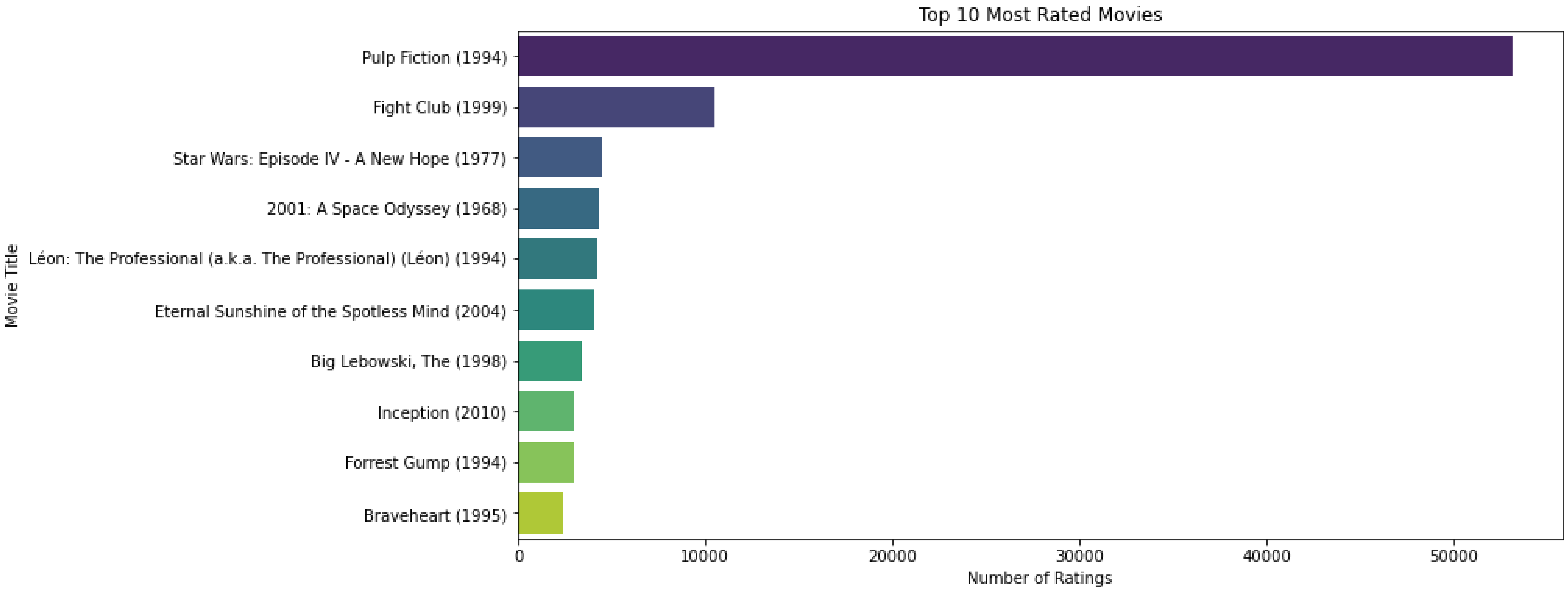
Sci-fi, thought-provoking and **twist ending** are the most common movie tags.

DISTRIBUTION OF TOP GENRES



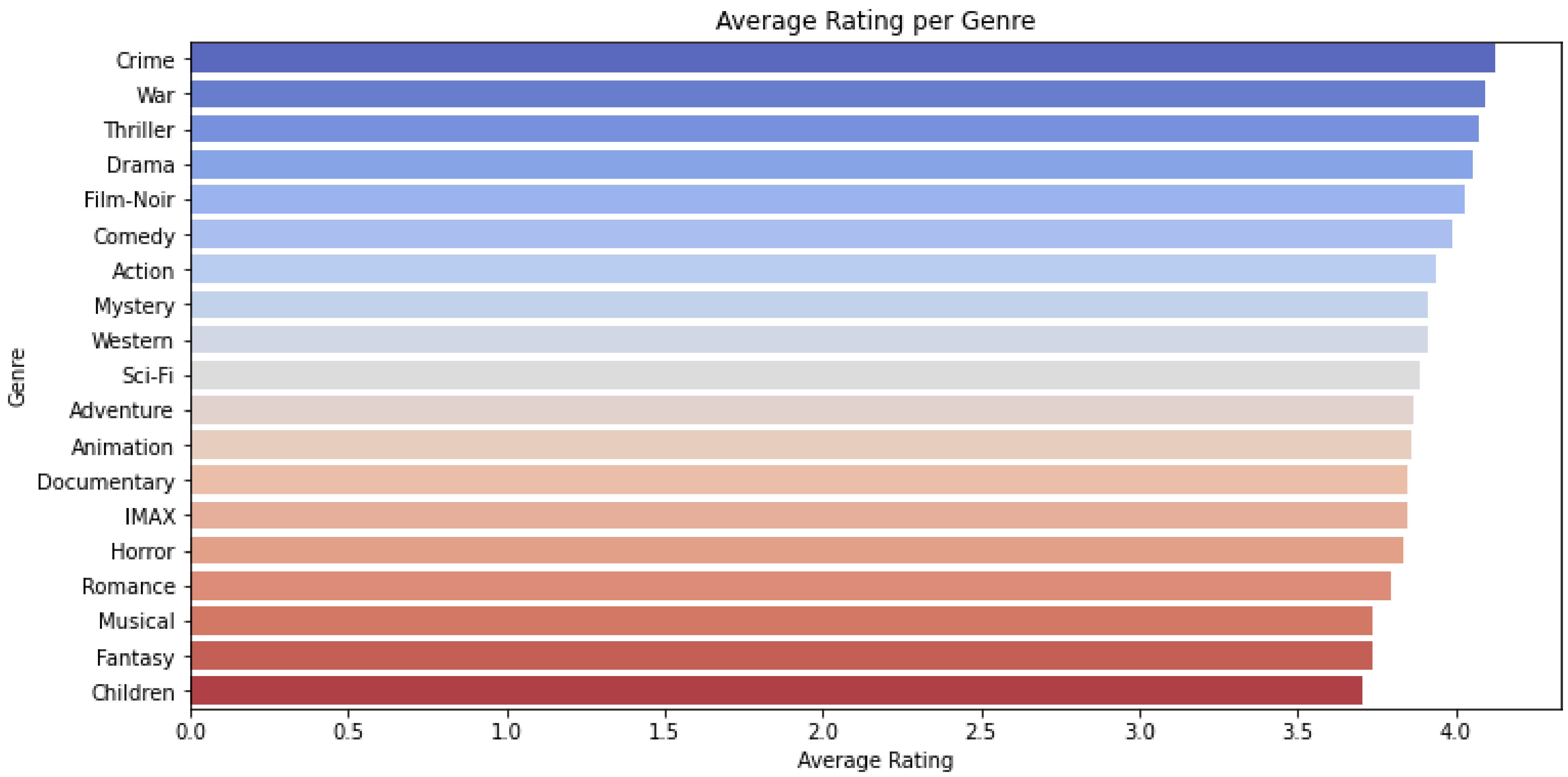
Drama(14000) is the top genre.

DISTRIBUTION OF RATINGS BY MOVIES



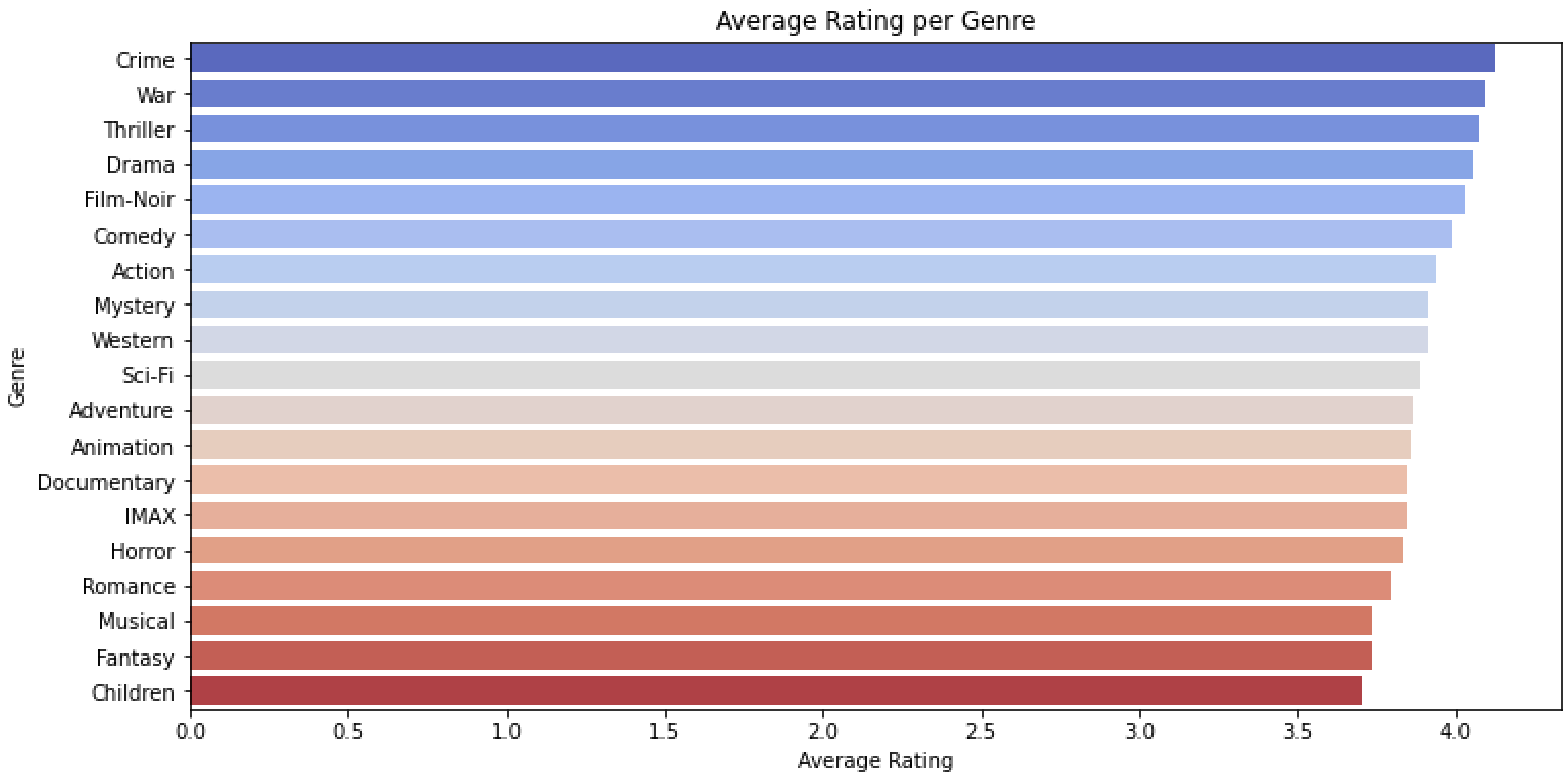
Pulp Fiction is the most rated movie.

DISTRIBUTION OF RATINGS BY GENRES



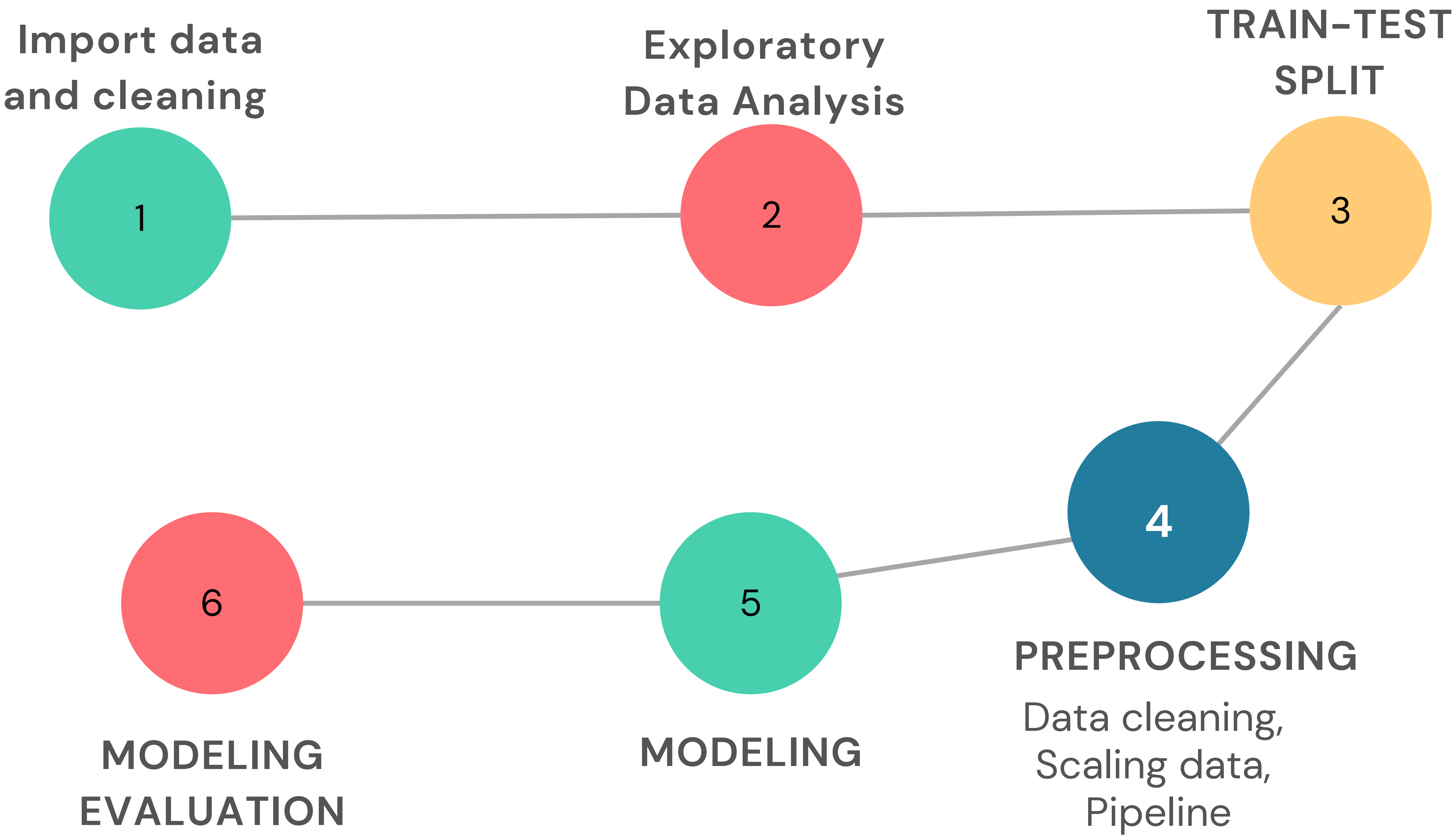
Crime | War | Thriller | Drama | Film-Noir - are the top rated genres, with an average of **4.0+**

DISTRIBUTION OF RATINGS BY GENRES



Crime | War | Thriller | Drama | Film-Noir - are the top rated genres, with an average of **4.0+**

MODELING


Import data and cleaning
1
Exploratory Data Analysis
2
TRAIN-TEST SPLIT
3
4
PREPROCESSING
Data cleaning,
Scaling data,
Pipeline
5
MODELING
6
MODELING EVALUATION

MODEL EVALUATION

KNNBasic (PEARSON Similarity)	RMSE	0.7246
	MAE	0.5155
KNN with Means	RMSE	0.6955
	MAE	0.4914
KNN Baseline (PEARSON Similarity)	RMSE	0.6147
	MAE	0.4171
SVD (Matrix Factorization)	RMSE	0.5081
	MAE	0.3467
SVD (with GridSearchCV)	RMSE	0.2465

MODEL EVALUATION

FINAL MODEL

SVD (with GridSearchCV)	RMSE	0.2465

The SVD model with tuned parameters provided the best performance, with the lowest RMSE metrics.

RECOMMENDATION

- Implement the collaborative filtering model on the platform to provide real-time recommendations.
- Regularly update the system with new user data for continuous improvement.
- Continuous Learning:Add new user ratings and movie information over time to improve model accuracy.
- Additional Data:Incorporate user feedback and social media interactions to further refine recommendations

CONCLUSION

- Collaborative Filtering proved to be the most accurate model for movie recommendations.
- Content-Based Filtering is useful for new users with fewer ratings.
- The recommendation system can significantly enhance user experience by providing personalized suggestions based on historical data.

NEXT STEPS

- Explore user segments (e.g., age groups, regions) to deliver more personalized recommendations.
- Address the issue of recommending for new users with no ratings by leveraging content-based filtering and hybrid approaches.
- A/B Testing: Continuously monitor and improve the system by conducting A/B tests on recommendation quality and user engagement.

THANK YOU