MOVIE RECOMMENDATION SYSTEM

MOVIEXPLOSION

OUR TEAM

Amos Kipkirui

Robert Mbau

Fiona Kungu

Maureen Kitanga

Dennis Mwanzia

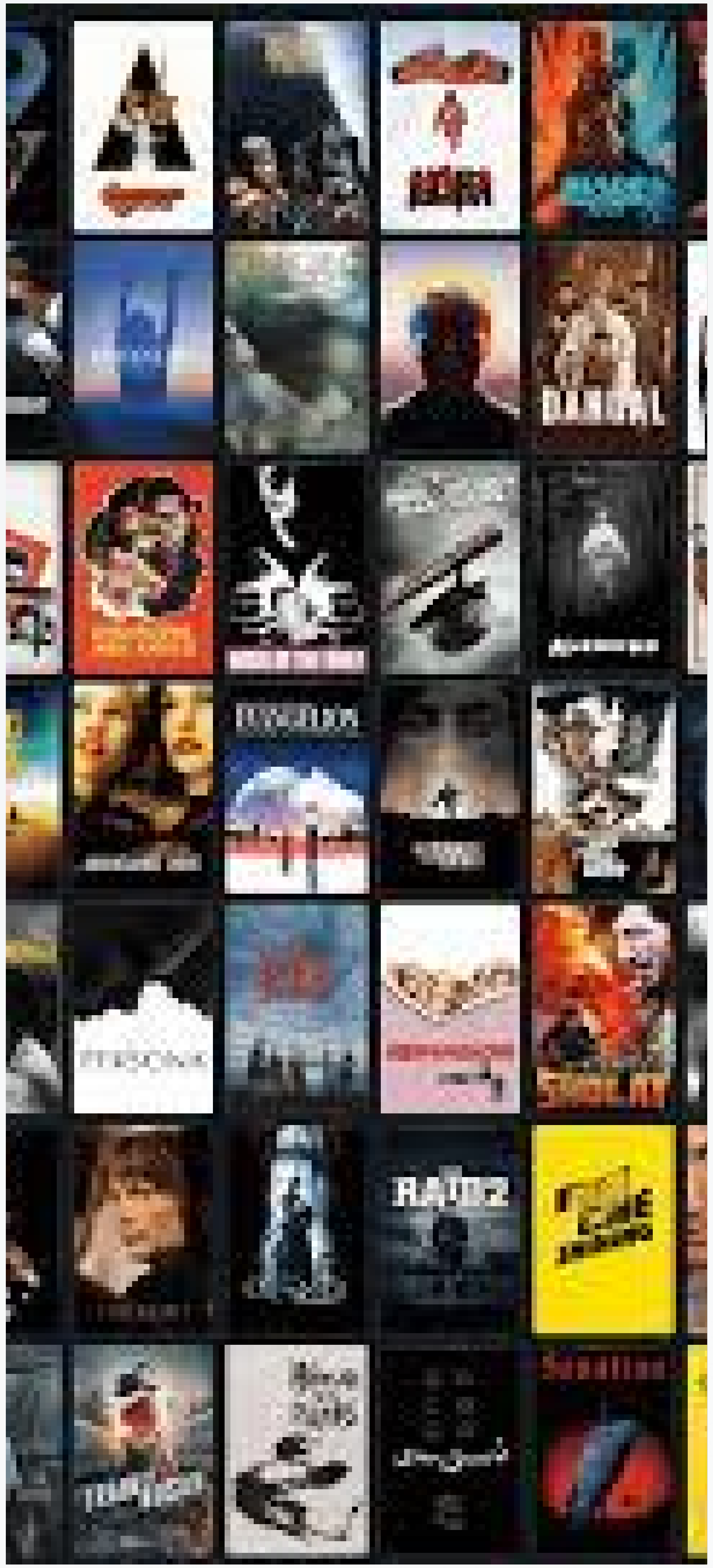
Edwin Muhia

OUTLINE

OVERVIEW

MovieXplosion, a new streaming platform wants to improve their user satisfaction. The performance of the platform is dependent on how they can keep users engaged, one way to do this is by providing tailor-made recommendations to the users and drive them to spend more time on the platform.

The project aims to develop a system that recommends movies to users based on their ratings of previous watched movies or preferred genre . We will implement this using collaborative filtering, content-based filtering and hybrid filtering approaches



PROBLEM STATEMENT

- Lack of personalized recommendations for new users- The current system fails to provide personalized recommendations to new users, which hampers their initial experience and engagement with the platform. Without tailored recommendations based on their preferences, new users may struggle to find relevant movies, leading to lower satisfaction and potentially discouraging them from further interact.

- Inability to offer customized recommendations to existing users: The existing users of the platform are not receiving tailor-made recommendations that align with their evolving preferences and interests. As a result, they may encounter movies that are irrelevant or less appealing to them. The lack of personalized recommendations reduces user satisfaction and diminishes the likelihood of user retention over time

DATA UNDERSTANDING

- The data used was sourced from [MovieLens] (https://grouplens.org/datasets/movielens/latest/),Even though we had a dataset of roughly 27 million, we used a small dataset of 100,000 rated and tagged movies, due to limited computational power. The data contains information about movies, ratings by users and other relevant information.

There are several files available with different columns:

1. Movies File

- It contains information about the movies.

2. Ratings file

- It contains the ratings for the movies by different users

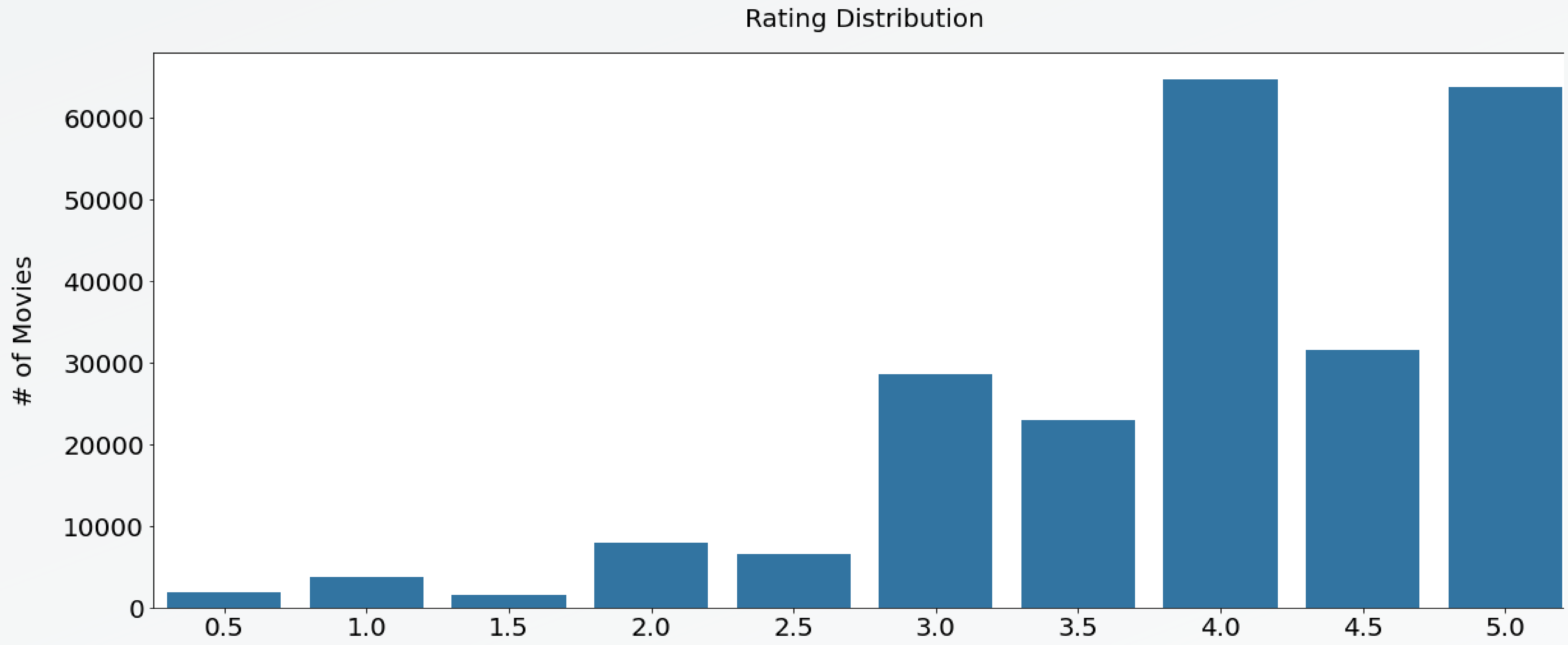
3. Tags file

- It has user-generated words or short phrases about a movie with the meaning or value being determined ny the specific user.

inks file

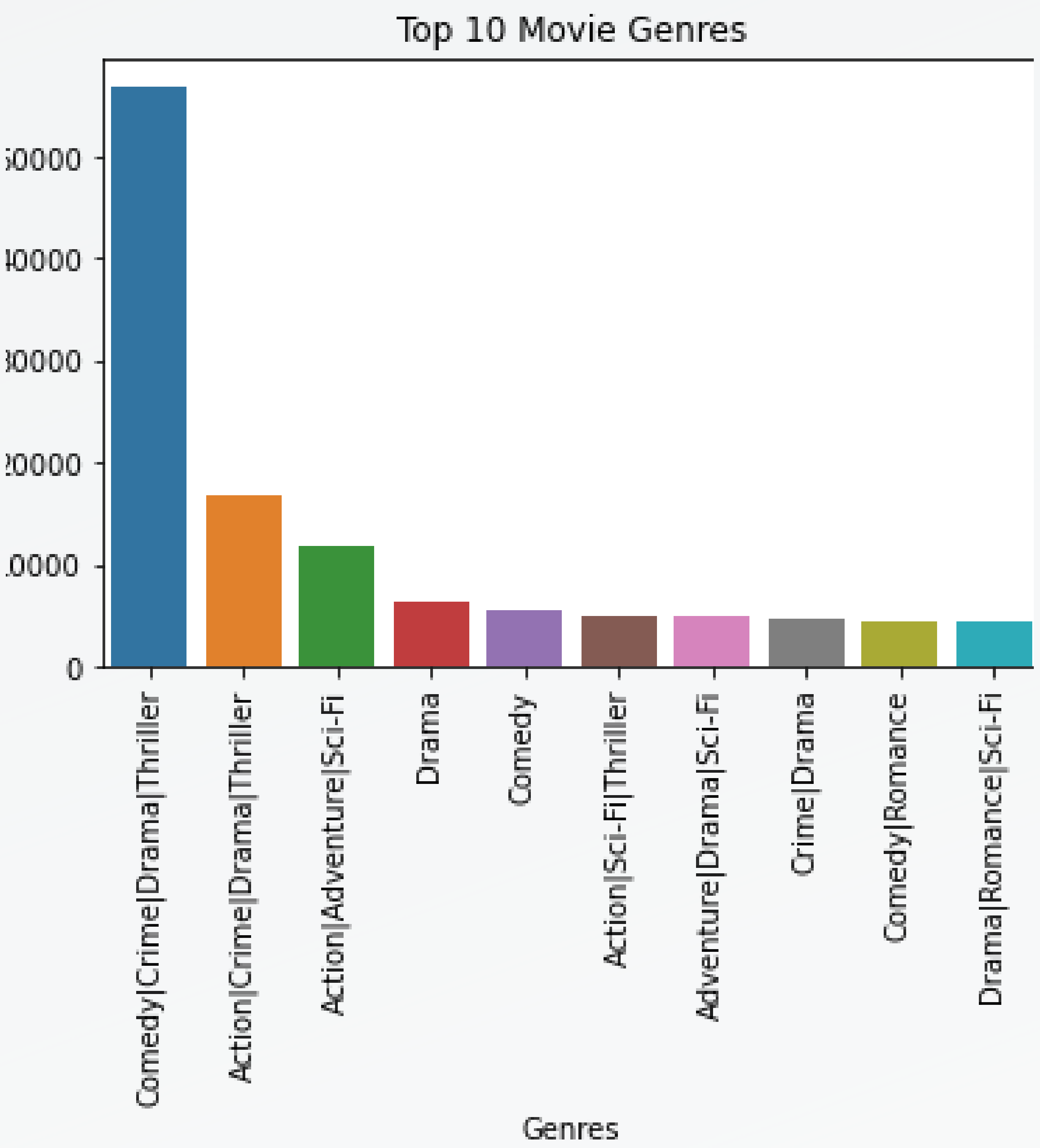
- This are identifiers that can be used to link to other sources of movie data as provide by MovieLens

EXPLORATORY DATA ANALYSIS

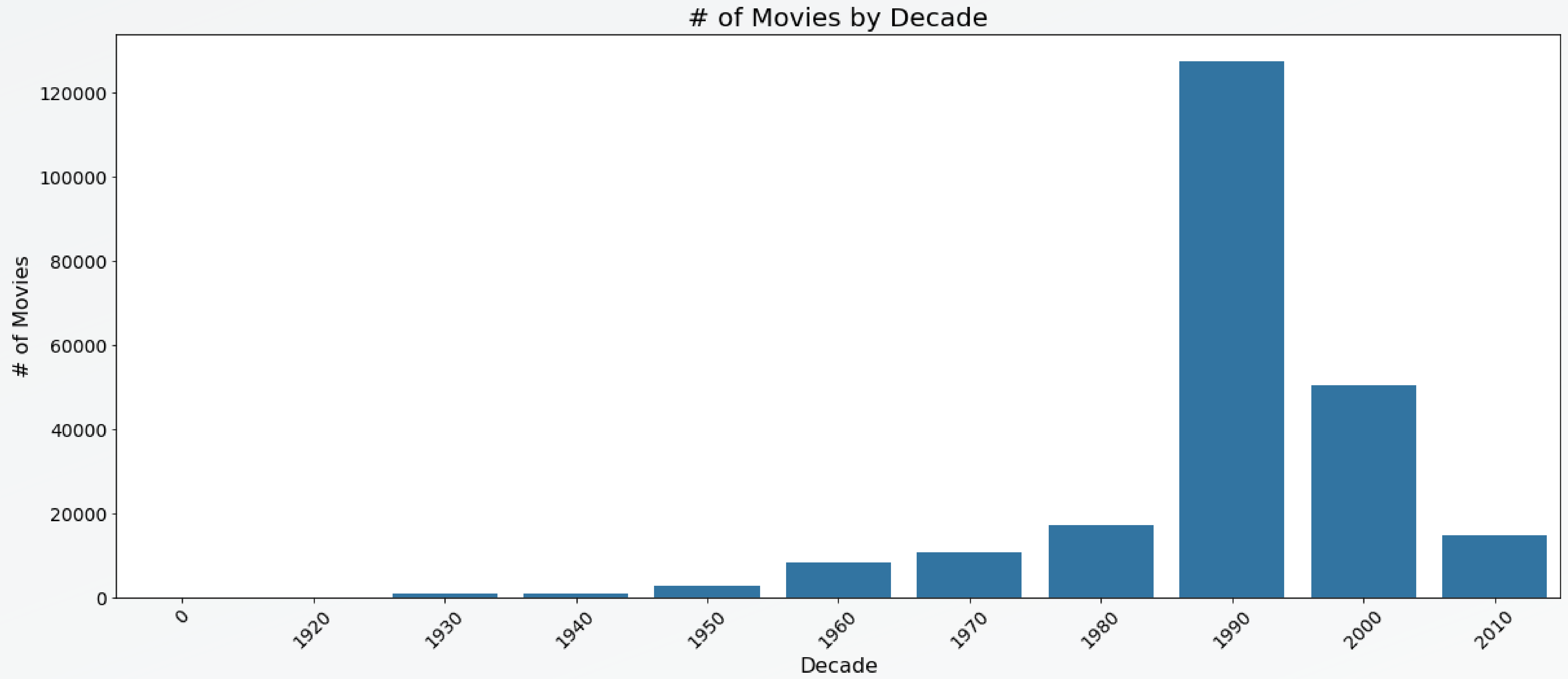


- The highest rated movies got `5` while the lowest rated was `0.5`.
- Most movies were rated at `5` and `4` with a few getting ratings of `0.5` and `1.5`

TOP TEN GENRE MOVIES

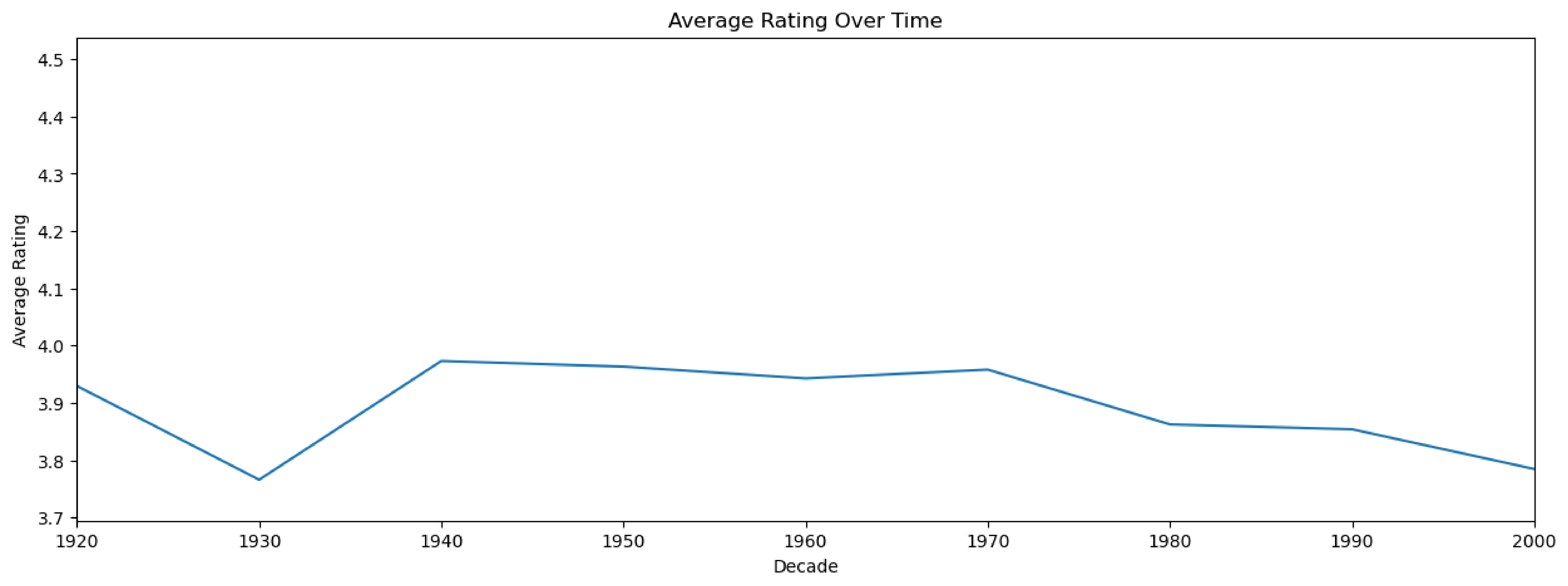


MOVIE PRODUCTION BY DECADE



- Movie production increase steadily from the 1920s and reached its peak in the 1990s before declining in the new millennium to the same levels as the 1980s.

AVERAGE RATING OVER TIME



- The lowest average rating was recorded for movies produced produced in 1930s while movies produced in 1940s got the highest average rating.
- Generally, average ratings have been declining since 1950s with a very sharp decline observed between 1970 and 1980.
- The year of movie production will be very useful to determine the average rating of a movie

MODELLING

In order to conduct this evaluation, we split our dataset into training and test sets, with a 20% test size, ensuring that our model's performance is assessed on unseen data. We then proceeded to evaluate three different models:

1. SVD (Singular Value Decomposition): This model leverages matrix factorization techniques to uncover latent factors and generate personalized recommendations.
2. KNNBasic with Pearson correlation: This model utilizes the K-nearest neighbors algorithm, considering the similarity between users based on Pearson correlation, to provide recommendations.
3. KNNWithMeans with Pearson correlation: Similar to KNNBasic, this model incorporates the mean ratings of users to improve the accuracy of recommendations.

For each model, we performed 5-fold cross-validation, measuring the Root Mean Squared Error (RMSE) as our evaluation metric. The RMSE quantifies the average difference between the predicted and actual ratings.

MODELLING

Model Comparison:

- SVD: RMSE = 0.8723
- KNNWithMeans: RMSE = 0.8971
- KNNBasic: RMSE = 0.9731

FINDINGS

- KNNBasic: RMSE = 0.973The SVD model achieved the lowest RMSE of 0.8723, indicating superior performance compared to the other models.
- KNNWithMeans had the second-best performance with an RMSE of 0.8971.
- KNNBasic had the highest RMSE of 0.9731, indicating the lowest level of accuracy among the evaluated models.

Best Performing Model:

- The SVD model demonstrated the highest accuracy in generating recommendations.
- With an RMSE of 0.8723, it outperformed both KNNWithMeans and KNNBasic.

BENEFITS AND RECCOMMENDATIONS

- By implementing the SVD model as our recommendation system, we can provide more accurate and reliable recommendations to our users.
- Improved accuracy leads to enhanced user satisfaction and engagement.
- Higher user satisfaction increases customer loyalty and positively impacts business outcomes
- Integrate the SVD model into our recommendation system.
- Continuously monitor and evaluate the system's performance.
- Explore further enhancements and optimizations to provide even more precise recommendations.

A WARM THANK YOU TO ALL OF YOU!