
MOVIE
RECOMMENDATION
SYSYTEM
STEAMFLIX

GROUP 8 MEMBERS

1. SAMUEL GICHURU
2. NASHON OKUMU
3. SHEILA MULWA

OUTLINE

01 OVERVIEW

02 PROBLEM STATEMENT
OBJECTIVES, CHALLENGES

03 DATA UNDERSTANDING

04 EXPLORATORY DATA ANALYSIS

05 MODELLING

06 EVALUATION

07 BENEFITS AND RECOMMENDATIONS

OVERVIEW

StreamFlix, a rapidly growing streaming platform that aims to improve their user satisfaction by enhancing engagement on the platform. A key strategy for achieving this goal is to offer personalized recommendations, encouraging users to discover more content and extend their viewing time.

The goal of this project is to create a system that recommends movies to users. We plan to achieve this by using collaborative filtering, content-based filtering, and hybrid methods.

PROBLEM STATEMENT

- The existing system on the platform fails to deliver adequate recommendations to users, resulting in low levels of engagement, satisfaction, and retention. Additionally, the system lacks the capability to offer quality recommendations to new users, and existing users do not receive personalized suggestions.

- The new system is designed to avoid these problems and offer relevant recommendations to every user.

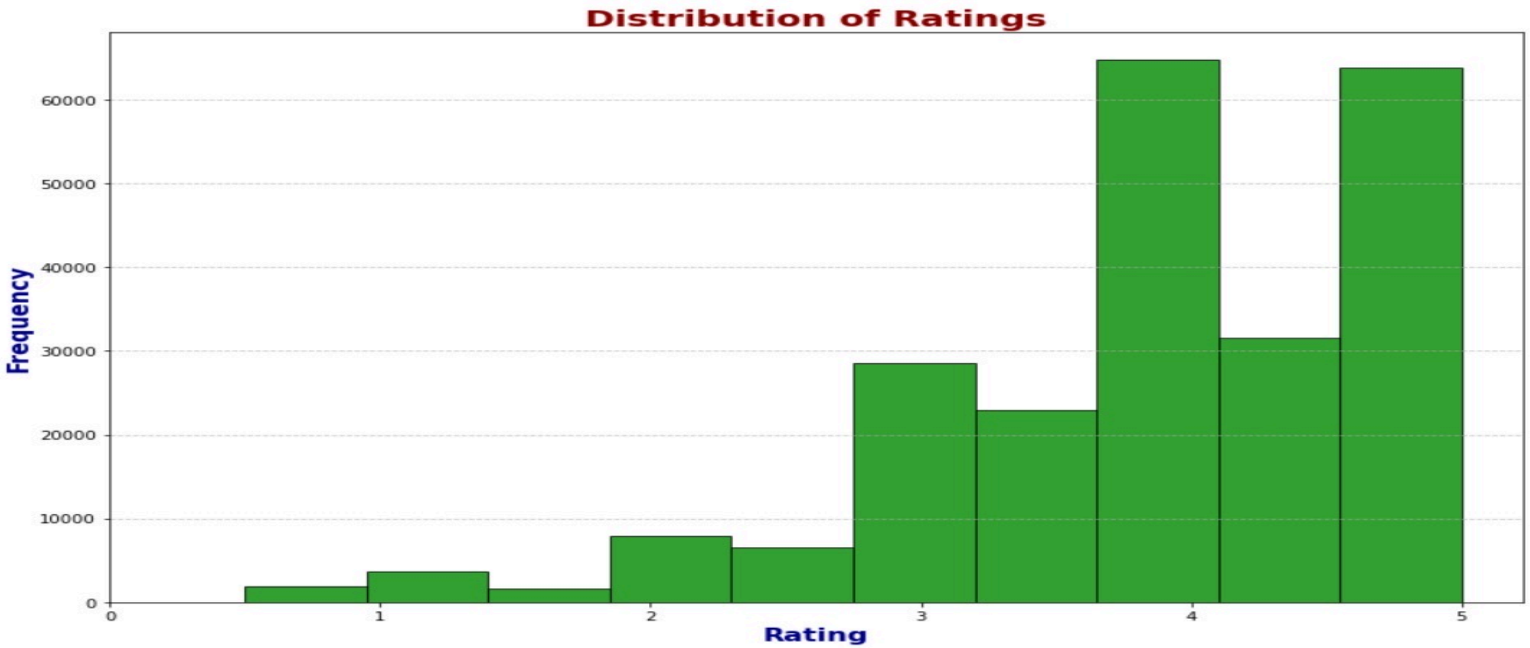
DATA UNDERSTANDING

- The data used was sourced from [Movie Lens]

(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 l imited computational power. The data contains information about movies, ratings by users and other relevant information.

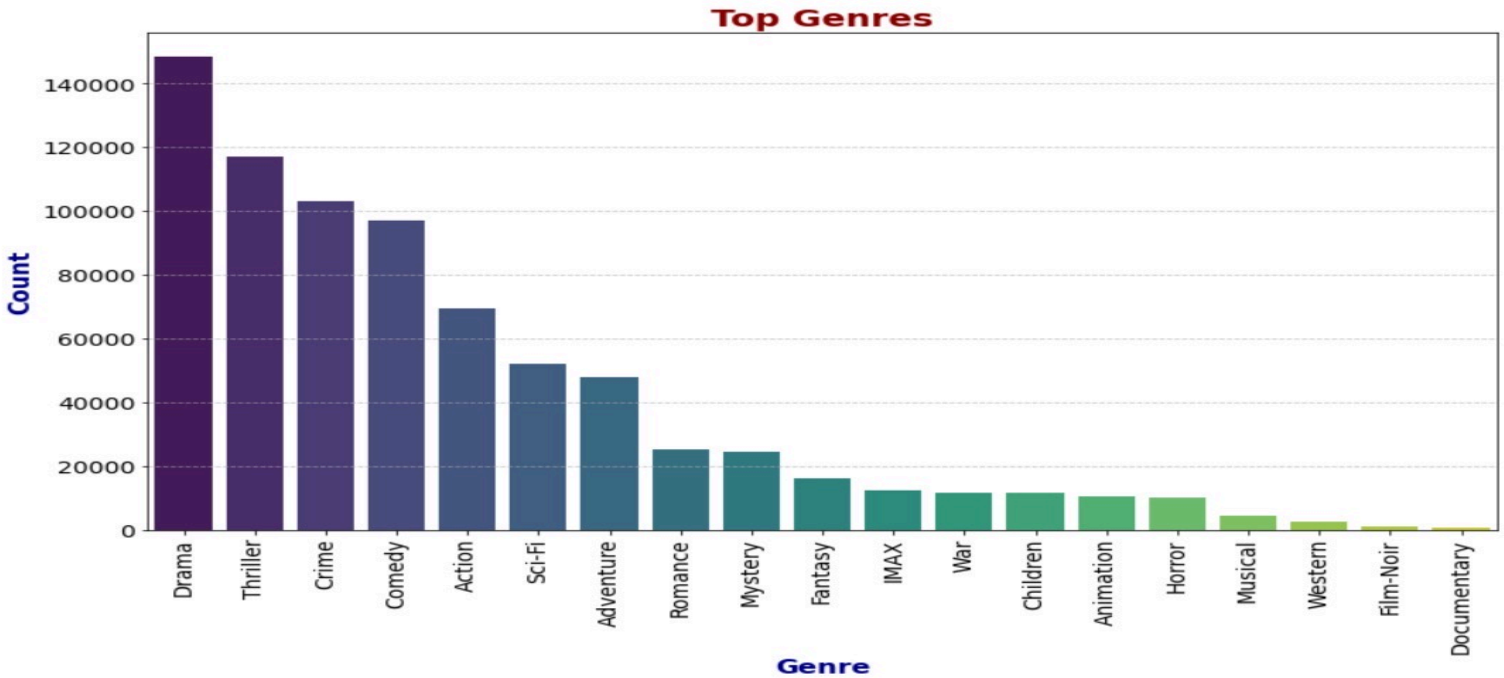
There are several f i les available with different columns: 1 . Movies File
I t contains information about the movies.
2 . Ratings f i le
I t contains the ratings for the movies by different users 3. Tags f i le
I t has user-generated words or short phrases about a movie with the meaning or value being determined ny the specific user.
inks f i le
This are identifiers that can be used to l ink to other sources of movie data as provide by Movie Lens

EXPLORATORY DATA ANALYSIS



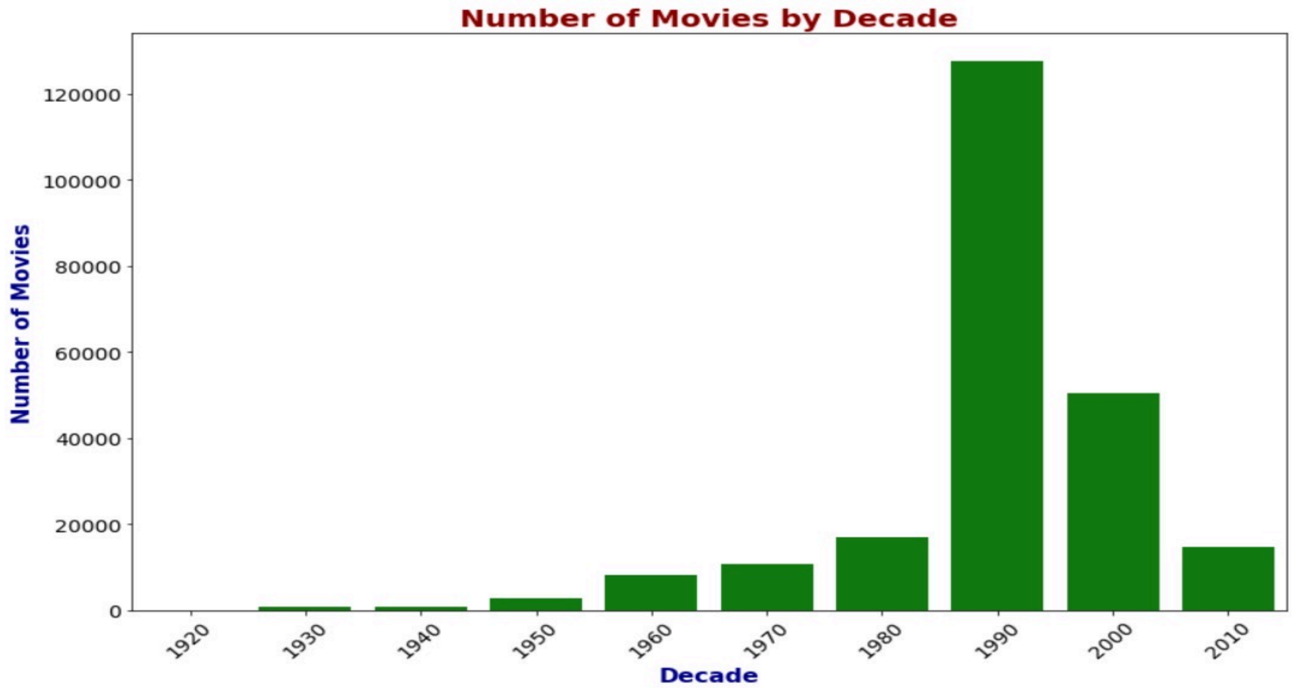
The majority of movie ratings are high, with 4 and 5 being the most common, indicating a positive user bias. Negative ratings (1 and 2) are significantly less frequent, suggesting users are less inclined to give poor ratings

TOP MOVIE GENRES



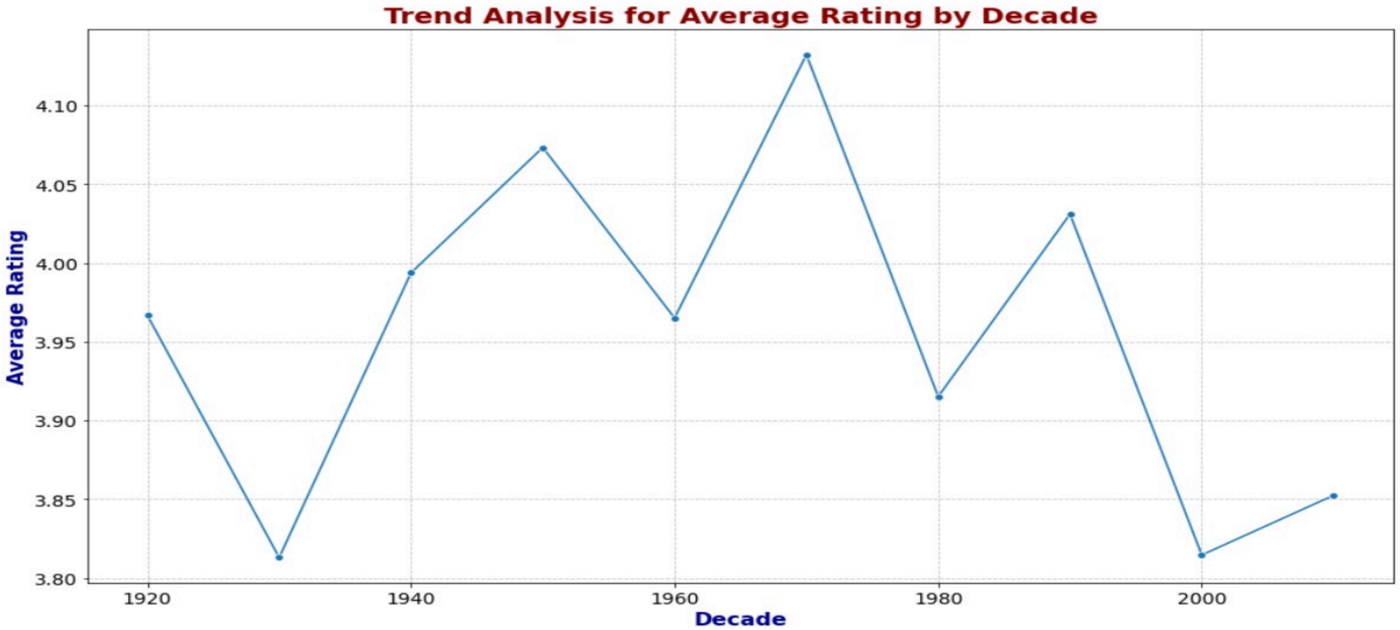
The "Comedy|Crime|Drama|Thriller" genre combination is the most popular, significantly outpacing others with over 50,000 counts. Other notable genre combinations include "Action|Crime|Drama|Thriller" and "Action|Adventure|Sci-Fi," indicating diverse audience preferences.

MOVIE PRODUCTION BY DECADE



The 1990s experienced a peak in movie production, with over 120,000 films released. There was a steady increase in the number of movies from the 1920s to the 1980s, reflecting the industry's growth. However, the number of movies released declined significantly in the 2000s and further in the 2010s.

AVERAGE RATING OVERTIME



- **The average ratings show a fluctuating trend with peaks in the 1940s and 1960s and notable declines in the 1930s, 1970s, and 1990s**
- **The 1960s marked the highest average rating. The lowest average rating was in the 1930s, followed closely by the 2000s.**

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:

SVD (Singular Value Decomposition): This model leverages matrix factorization techniques to uncover latent factors and generate personalized recommendations.

KNNBasic with Pearson correlation: This model utilizes the K-nearest neighbors algorithm, considering the similarity between users based on Pearson correlation, to provide recommendations.

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 EVALUATION

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

•**Model Enhancements:** To further refine the recommendation system, consider integrating advanced algorithms such as deep learning-based models and improved matrix factorization techniques. These enhancements can provide more accurate and nuanced recommendations, enhancing user satisfaction.

•**Data Expansion:** Expand the dataset to include a wider variety of movies, user demographics, and additional ratings. A richer dataset will improve the system's ability to capture diverse user preferences and deliver more relevant recommendations

•**User Interface Improvements:** Enhance the user interface by incorporating features like movie previews, user reviews, and genre-specific filters. These additions will make the system more engaging and useful, providing users with a richer browsing experience.

•**Scalable Deployment:** Deploy the system on a scalable cloud platform to handle increased user traffic and ensure consistent performance. This will support growth and provide a reliable user experience as the platform expands.


THANK YOU