RECOMMENDATION SYSTEMS

***BY** Gideon Ochieng, Lorna Wangui, Ann Mwangi,*
Charles Odhiambo,Victor Masinde

Introduction

- Our hybrid recommendation system offers personalized movie suggestions, enhancing user experience on streaming platforms.
- It combines collaborative filtering and content-based filtering to deliver accurate recommendations.
- The system improves user engagement and satisfaction.
- It provides valuable insights into audience preferences.
- Benefits include enhanced experiences for end users, improved engagement for streaming platforms, and valuable data for studios.

DATASET OVERVIEW

Movies DataFrame:

- **Entries:** 9,742
- **Columns:** movieId, title, genres

Ratings DataFrame:

- **Entries:** 100,836
- **Columns:** userId, movieId, rating

Tags DataFrame:

- **Entries:** 3,683
- **Columns:** userId, movieId, tag

Links DataFrame :

- **Entries:** 9,742
- **Columns:** movieId, imdbId, tmdbId

DATA CLEANING AND PREPROCESSING

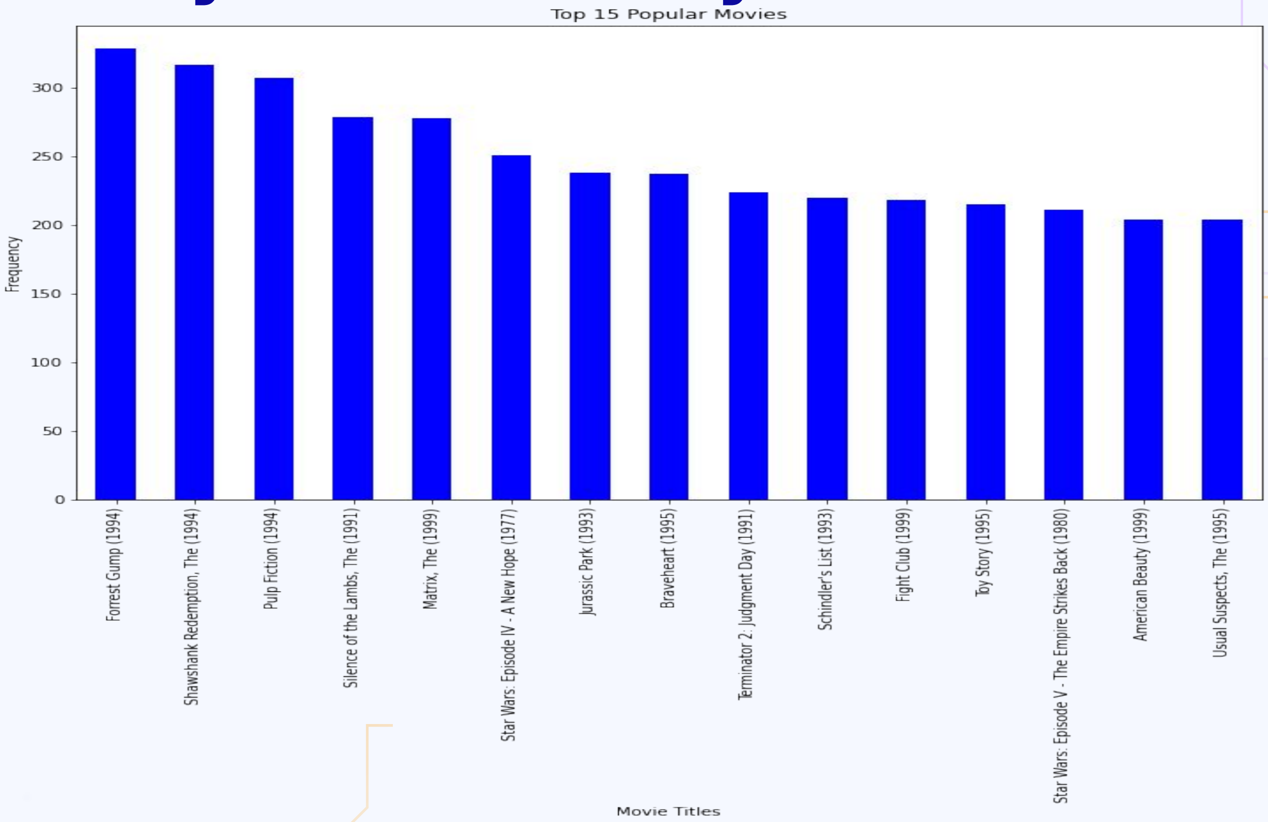
- The datasets had no missing values and duplicated rows

Merging Movies and Ratings Datasets

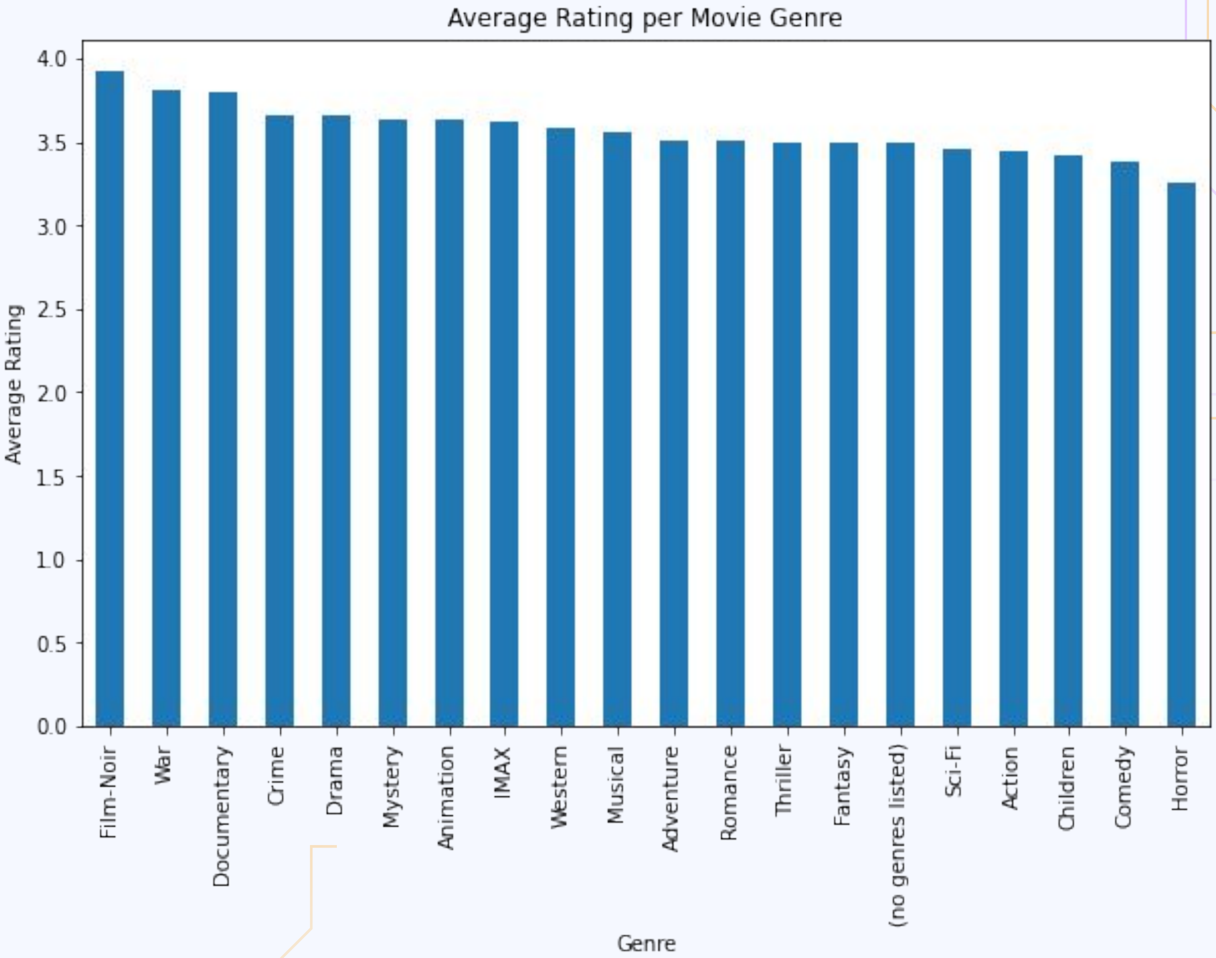
- The movies_df and ratings_df Data Frames are merged using the movieId column as the key.

Exploratory Data Analysis

The bar chart titled "Top 15 Popular Movies" shows that **Forrest Gump (1994)** has the highest frequency of mentions, while **The Usual Suspects (1995)** has the lowest among the top 15. This highlights the most popular movies based on frequency, providing insights into audience preferences.

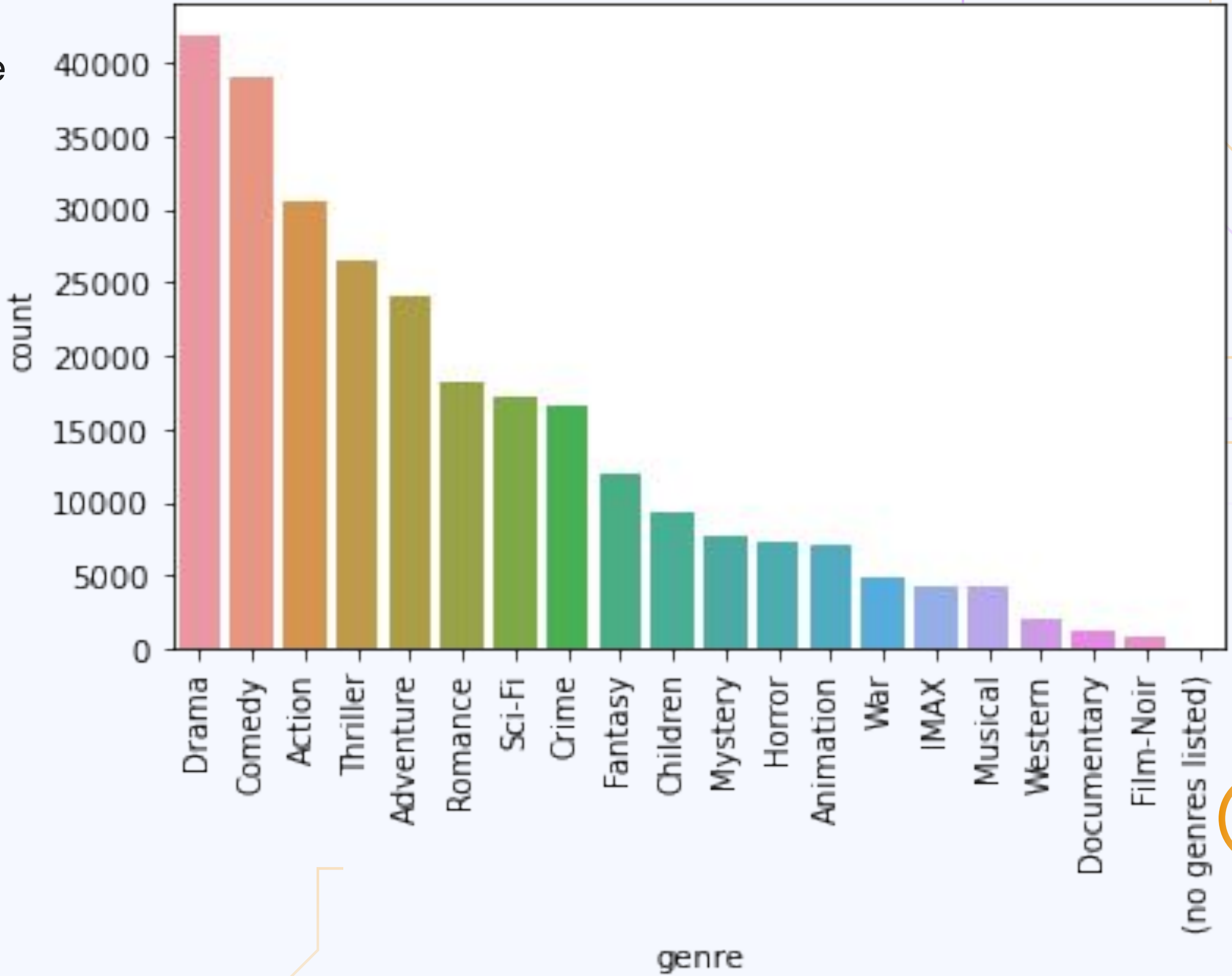


The bar chart titled "Average Rating per Movie Genre" shows that **Film-Noir** has the highest average rating, while **Horror** has the lowest. It highlights the varying average ratings across different movie genres, providing insights into audience preferences.

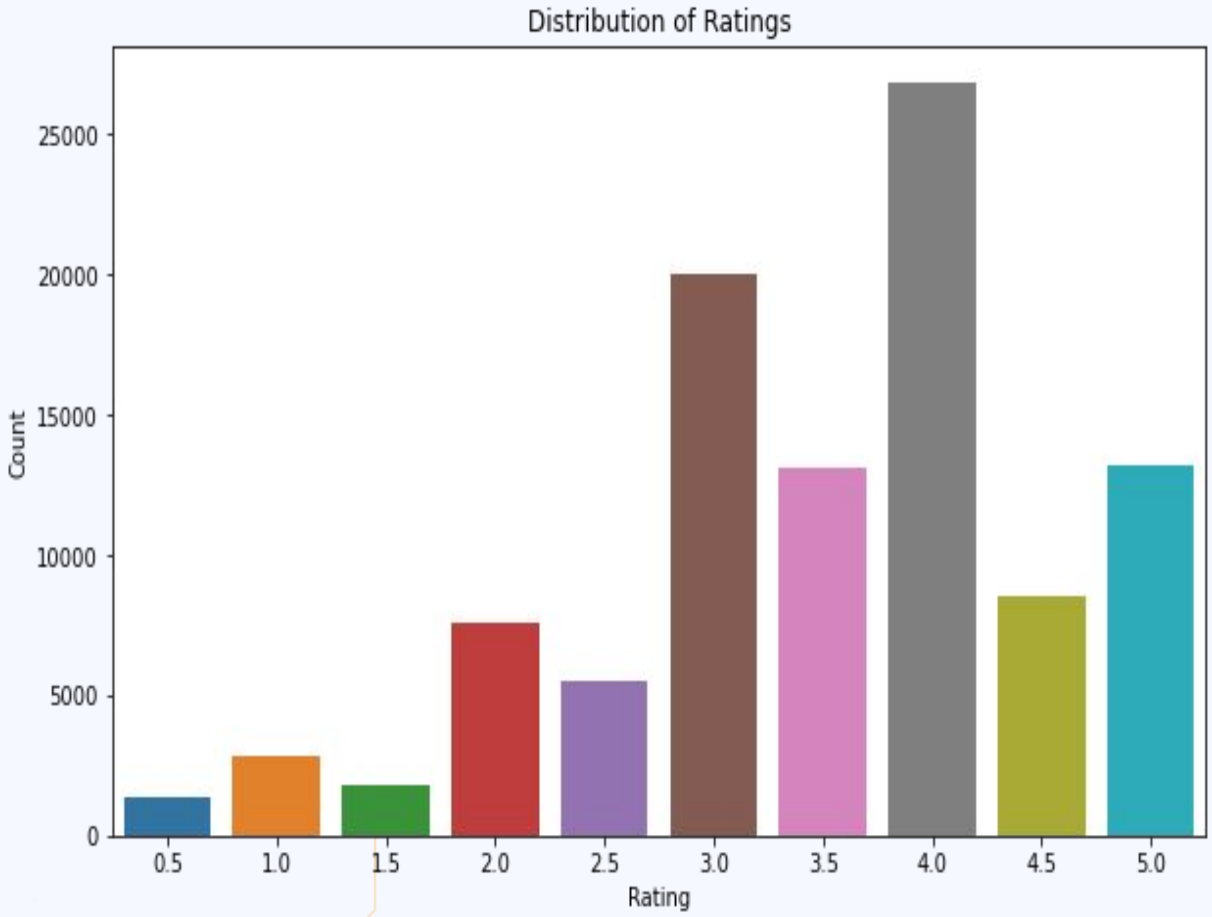


- Dominant genres like Drama and Comedy should be prioritized in recommendations, as they are most popular.

- Recognizing lesser-represented genres can enhance diversity and personalization in recommendations.



- The bar chart above shows the distribution of ratings in the dataset.
- This distribution suggests a positive skew, with higher ratings being more common.



PREPARING DATA

Filter Relevant Columns

Define Rating Scale

Load Data into Surprise – we used `Dataset.load_from_df` to format the data for collaborative filtering.

Divide data into training (80%) and testing (20%) sets using `train_test_split`.

MODELLING

COLLABORATIVE FILTERING USING SVD

- Employs Singular Value Decomposition (SVD) to predict user ratings based on historical data.
- Technique that relies on user-item interactions to predict user preferences.

SVD Model:

- Matrix factorization technique that decomposes the user-item interaction matrix into latent factors.

Performance Metrics:

- Evaluated using RMSE and MAE:
    - **RMSE:** 0.8688
    - **MAE:** 0.6681

CONTENT-BASED FILTERING USING COSINE SIMILARITY

- Recommends movies based on similarities in content
- Computing the cosine similarity matrix based on the TF-IDF vectors.

HYBRID RECOMMENDATION SYSTEM

- Combines collaborative filtering (SVD) and content-based filtering (cosine similarity) to provide more accurate and diverse recommendations.
- Generates recommendations by leveraging both user interaction data and content similarity.

RESULTS AND PERFORMANCE

Collaborative Filtering (SVD):

- RMSE: 0.8688, MAE: 0.6681

KNN Model:

- RMSE: 0.9414, MAE: 0.7193

Hyperparameter Tuning:

- Best RMSE score for SVD: 0.866918163661489
- Best RMSE score for KNNBasic: 0.9311428065101948

CONCLUSIONS

- The hybrid system demonstrates exceptional accuracy (RMSE: 0.8688) in predicting user ratings, ensuring personalized recommendations.
- The hybrid approach addresses the cold-start issue by using content-based filtering for new users, ensuring everyone receives valuable recommendations.
- By offering personalized and timely recommendations, the system boosts user engagement and satisfaction, leading to higher retention.
- The system is designed for continuous improvement with additional features, real-time updates, and user feedback integration

RECOMMENDATIONS

- Incorporate a wider range of movies, genres, and user interactions to improve accuracy and cater to diverse preferences.
- Regularly re-tune SVD and KNN model hyperparameters with GridSearchCV to ensure optimal performance.
- Allow users to rate and provide feedback on recommendations to continuously refine the algorithms.

Thank You!

Name	Github
Victor Masinde	@Masinde10
Gideon Ochieng	@OchiengGideon
Ann Mwangi	@ann-mw
Lorna Wangui	@lorna-creator
Charles Odhiambo	@T-hoveen