Movie Recommendation System Analysis

Data Exploration, Analysis, and Recommendation Algorithms

Group 8

1. David Mwai Gathimba
2. Hannah Nduta Karanja
3. Joan Owuor
4. Linda Ng’eno
5. Victor Muuo
6. Peter Njugu Murage

Business Understanding

In the competitive landscape of streaming services with numerous online platforms offering vast libraries of content, a personalized User Experience is crucial to deliver a tailored movie-watching experience which encourages longer viewing times, reduces churn rates and improves user satisfaction fostering a deeper connection between the user and the platform.

Stakeholders

Content Curators and Media Analysts

Product Managers

Marketing and Customer Success Teams

Problem Statement

Challenge

The increasing volume of movies available across various platforms, users often face difficulty in discovering movies that align with their preferences.

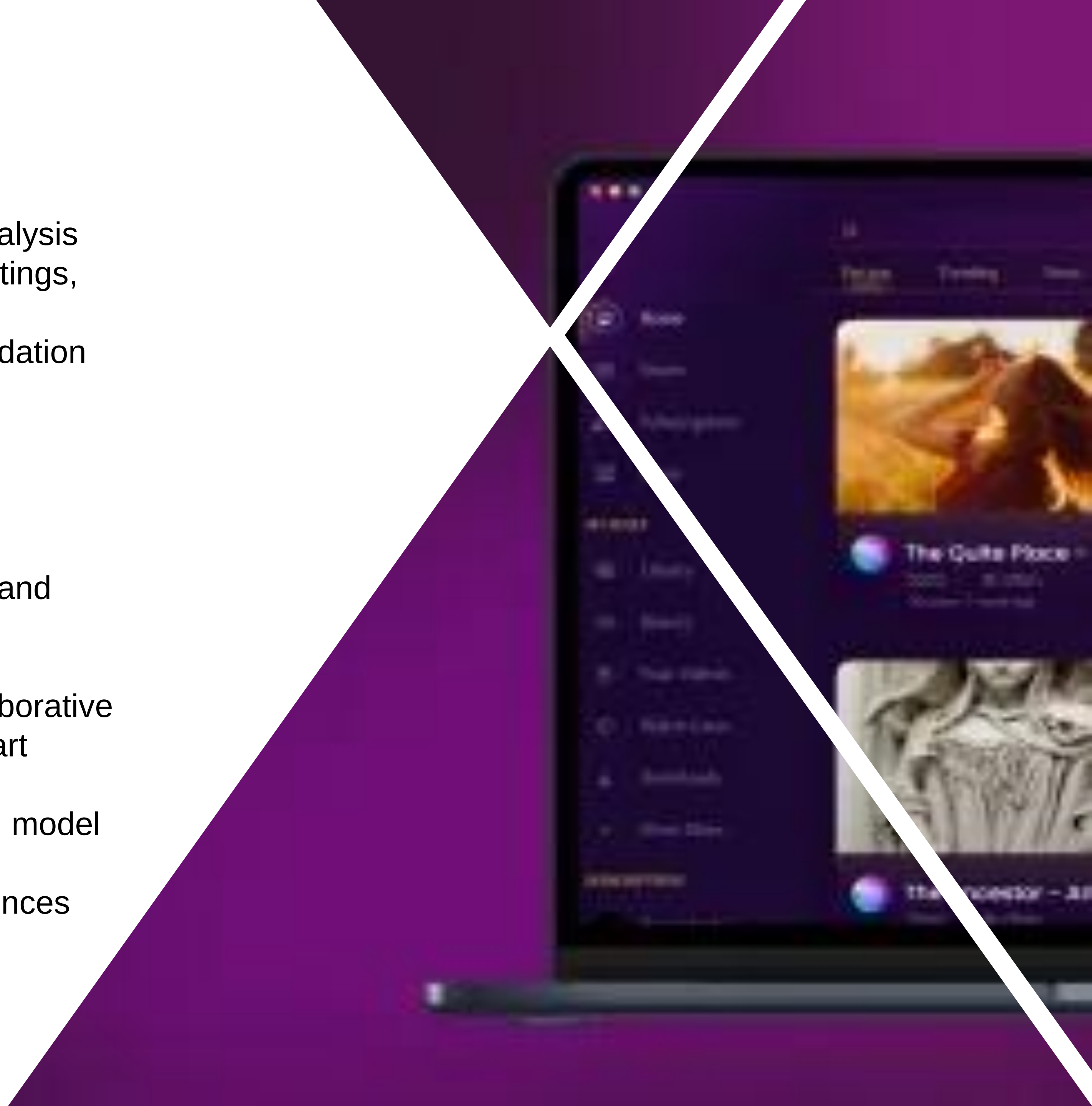
Goal

The goal is to develop an effective recommendation system and gain insights into user preferences and movie popularity.

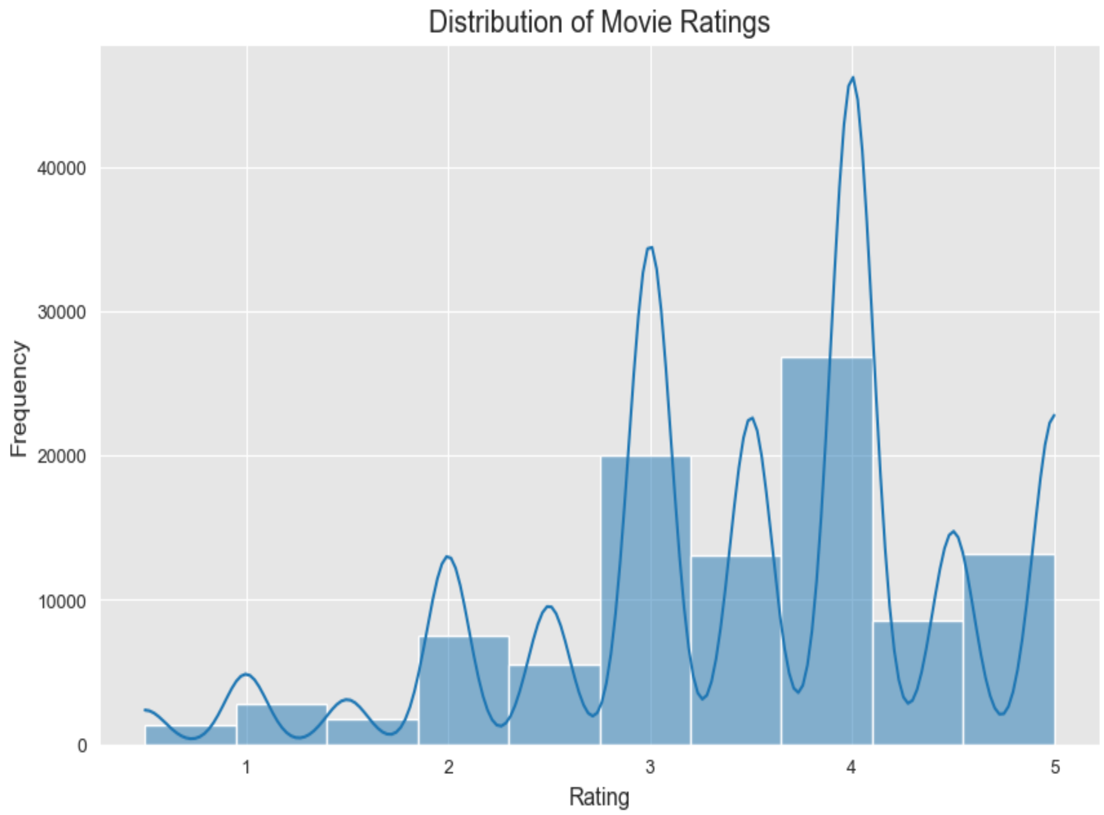
The system should cater to diverse user preferences and handle the cold start problem for new users with limited rating history.

Project Objectives

- To perform data cleaning and exploratory data analysis (EDA) on a movie dataset, which includes user ratings, movie metadata, and tags.
- To build and evaluate different filtering recommendation algorithms, including:
  - **User-based collaborative filtering**
  - **Item-based collaborative filtering**
  - **Model-based collaborative filtering**
  - **Content-based filtering**
- To compare the performance of these algorithms and identify the most effective approach for movie recommendations.
- Implement a hybrid approach that combines collaborative and content-based filtering to address the cold start problem.
- Evaluate the performance of the recommendation model using metrics like RMSE, MAE.
- Provide actionable insights based on user preferences and feedback for continuous improvement of the recommendation system.



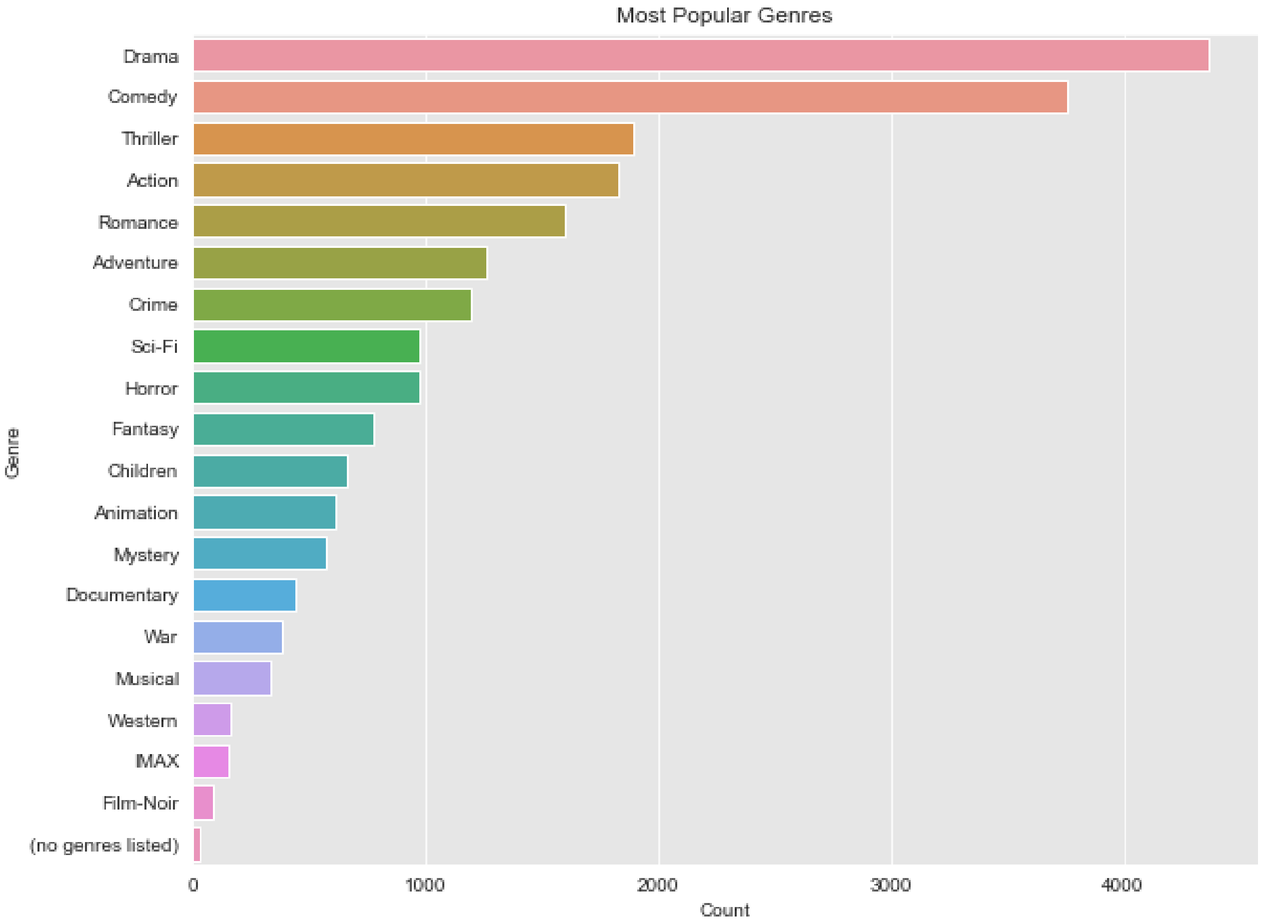
Visualizations: Distribution of Movie Ratings



The ratings are highly concentrated around 4, with a peak between 3.5 and 4. This suggests that users generally tend to rate movies favorably, with 4 being the most frequent rating.

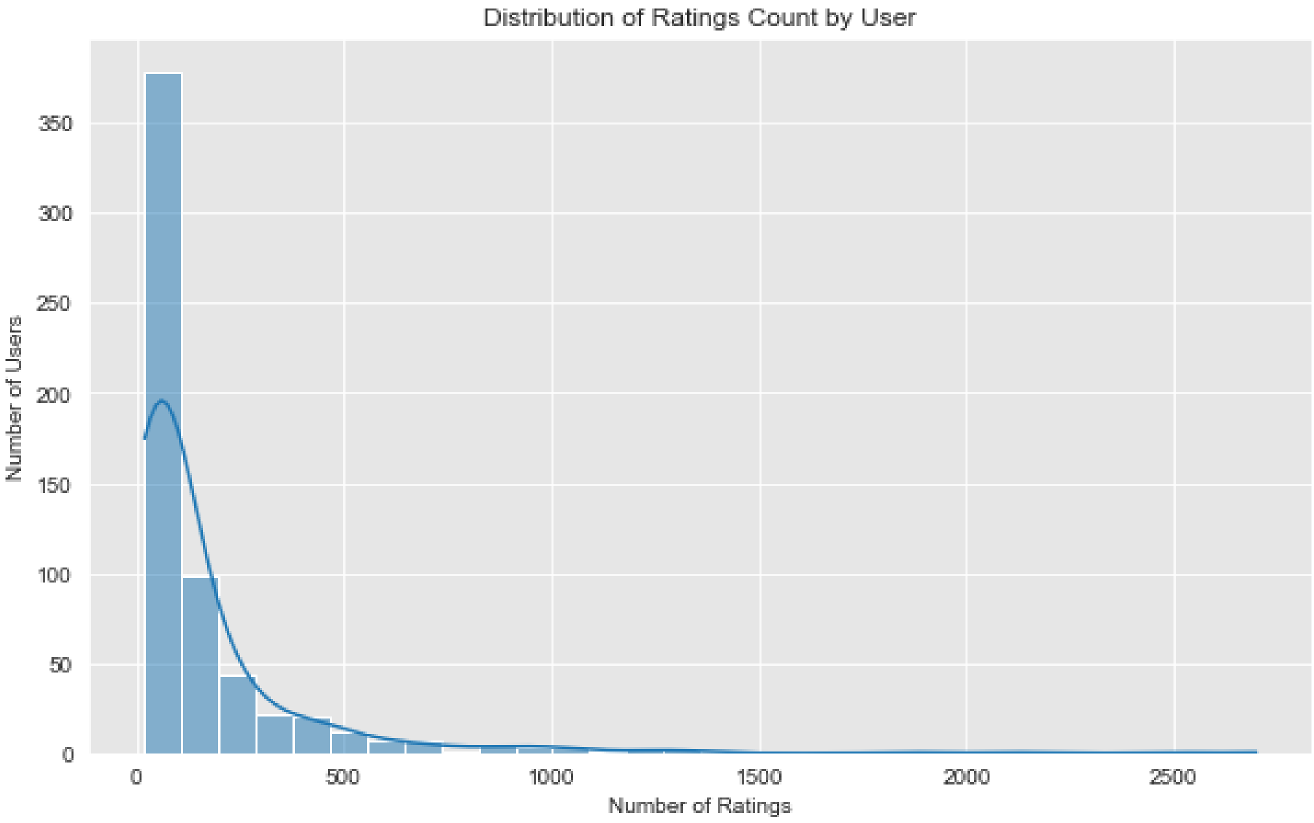
There is a clear tendency for users to give high ratings (between 3 and 5). Low ratings (below 2) are much less frequent, indicating that most movies are either liked or considered average by the users.

Visualizations: Most Popular Genres



`Drama` and `Comedy` are by far the most popular genres, with both appearing in over 4000 movies.
`Thriller`, `Action`, and `Romance` are also popular, though with slightly fewer appearances.
Genres like `War`, `Musical`, `Western`, and `IMAX` are much less frequent.
There's a small portion of movies where no genres are listed, which might be worth investigating.

Visualizations: Distribution of Rating Count by User



This visualization tracks how user ratings change over time. It's helpful for spotting trends, such as whether ratings for certain movies or genres are rising or falling in popularity.

Project Findings

Exploratory Data Analysis (EDA)

The dataset requires timestamp conversion. Ratings data reveals a distribution skewed towards higher ratings, with 4 and 5-star ratings being the most common.

Genre Distribution

Some genres are more popular than others, with Drama, Comedy, and Thriller being among the top genres.

Collaborative Filtering (User and Item-Based)

Collaborative filtering based on users and items provides recommendations based on similar users or movies. Both user-based and item-based collaborative filtering were implemented using k-nearest neighbors (KNN) and evaluated using metrics such as RMSE and MAE.

Model-Based Collaborative Filtering

Using Singular Value Decomposition (SVD), the system predicts ratings for unrated movies based on learned patterns in the ratings matrix. The SVD model yielded relatively low RMSE, indicating strong predictive performance.

Project Findings

Content-Based Filtering

By utilizing TF-IDF vectors of genres, this approach successfully recommended movies with similar genres to the target movie.

Hybrid Recommendation System

Combining collaborative filtering and content-based filtering with weighted scores resulted in a comprehensive approach that capitalizes on the strengths of both methods.

Performance Metrics

The RMSE and MAE metrics indicate that the model-based (SVD) and hybrid approaches provide a balance between accuracy and computational efficiency.

Limitations

Limitations

Cold Start Problem

New users with few or no ratings may not receive optimal recommendations due to insufficient data for accurate predictions.

Data Sparsity

The user-item matrix can be sparse, leading to challenges in finding similar users or items, which can affect recommendation accuracy.

Static Dataset

The MovieLens dataset may not reflect current trends, preferences, or newly released movies, limiting the system's ability to recommend the latest content.

Scalability

The collaborative filtering approach may face performance issues as the user base and item catalog grow, potentially slowing down recommendation generation.

Subjectivity of Ratings

User ratings are subjective and can vary widely; different users may interpret the rating scale differently, leading to inconsistencies in the dataset.

Implementation Complexity

While the hybrid approach can improve recommendations, it adds complexity to the system, requiring careful tuning and integration of multiple algorithms.

Recommendations

01
For production, implement a hybrid model with a tunable weight between collaborative and content-based scores to personalize recommendations further based on user behavior. This helps address the "cold start" problem for new users or movies with few ratings.

02
Implement strategies to account for user biases in ratings, such as users who consistently rate movies higher or lower than average. Normalizing ratings or weighing ratings based on user behavior can improve recommendation

03
Regularly update the recommendation system with new ratings and movie data to keep the recommendations relevant.

04
Monitor the system performance over time and retrain models periodically to maintain high accuracy as user preferences evolve.

THANK YOU

Movie Recommendation System Analysis