Movie Recommendation System

Members:

- ❑ Calistus Mwonga
- ❑ Brian Kanyenje
- ❑ Samwel Kipkemboi
- ❑ Kelvin Mutua
- ❑ Hannah Nyambura

OVERVIEW

This project aims to build a personalized movie recommendation system using the MovieLens dataset. It combines collaborative filtering and content-based filtering to produce movie recommendations for users. The system is designed to handle both regular users and new users with little or no rating history.

OUTLINE

- Business problem
- Business problem
- Objectives
- Data overview
- Visualizations and Results
- Business recommendations
- Conclusions
- Contacts

BUSINESS PROBLEM

Users on the StreamTimeNow platform often struggle to discover content that matches their preferences due to the vast number of available options. With thousands of movies to choose from, users are often overwhelmed and find it difficult to locate movies that match their interests, especially since most don't scroll deeply or explore the site extensively. This results in user-frustration, decision fatigue and in some cases platform abandonment.

As a result, StreamTimeNow faces a critical challenge in retaining users and maintaining long-term engagement, which directly impacts business sustainability.

OBJECTIVES

1

Identify Popular Movies

Identify the most popular movies based on metrics like number of ratings and average rating

2

Investigate User Ratings

Investigate how users rate movies

3

Develop Recommender Models

Developing and evaluating advanced recommender models (e.g., collaborative filtering, matrix factorization, or hybrid methods).

4

Explain Recommendations

Explaining recommendation results, using example user profiles and insights derived from the model.

DATA OVERVIEW

Dataset Source: The MovieLens dataset, specifically the 'ml-latest' version.

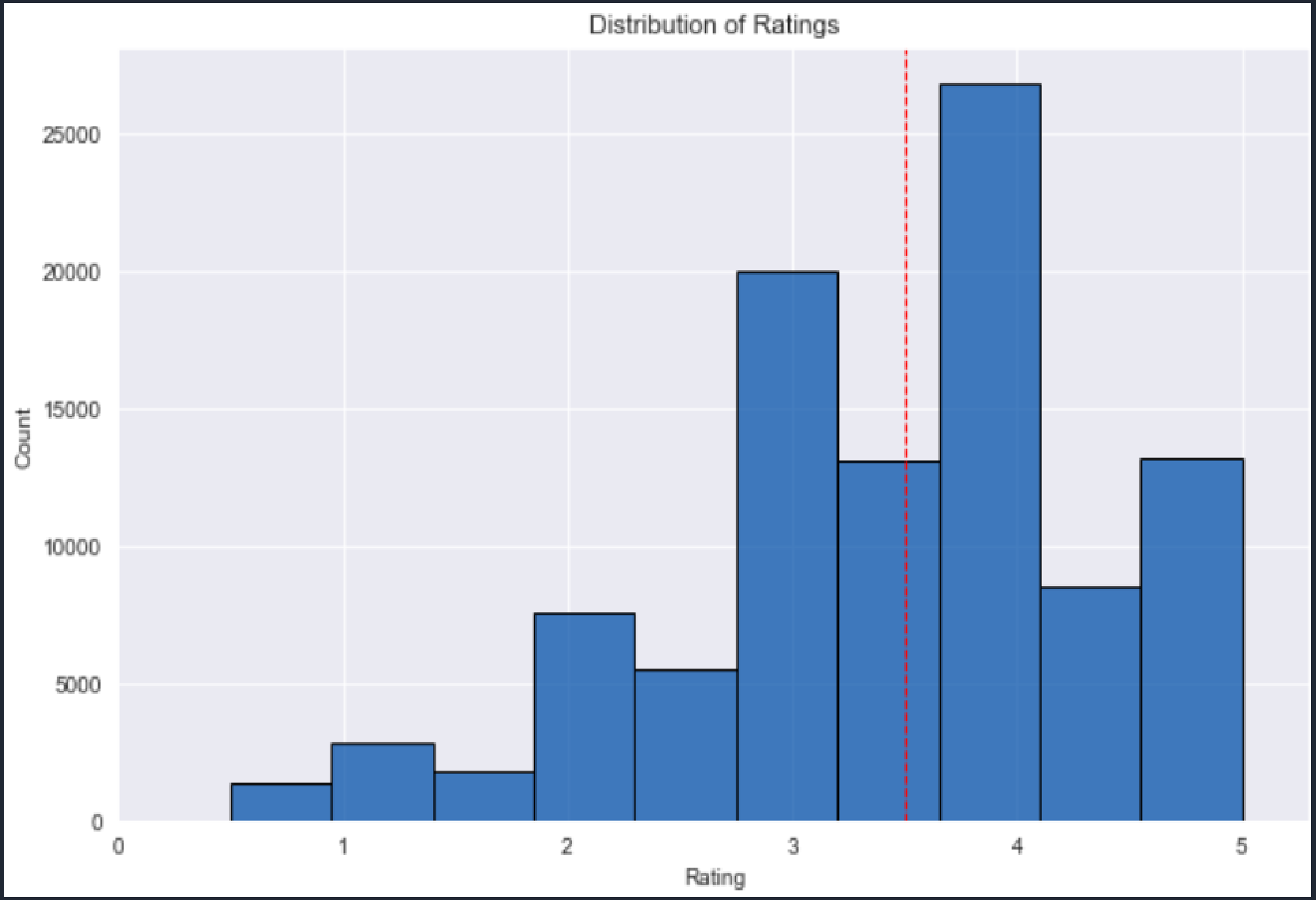
The dataset is quite large, containing over 100,000 ratings and information on nearly 10,000 movies.

What's Included: This dataset contains several files that give us different pieces of information:

- **ratings.csv:** This is the core file, containing user ratings for movies (User ID, Movie ID, Rating, Timestamp).
- **movies.csv:** This file provides details about the movies themselves (Movie ID, Title, Genres).
- **tags.csv:** This file includes user-generated keywords or tags applied to movies (User ID, Movie ID, Tag, Timestamp).
- **links.csv:** This file links the MovieLens IDs to external databases like IMDb and TMDb (Movie ID, IMDb ID, TMDb ID).

The key connection between all these files is the movieId. This allows us to link user ratings, movie details, user tags, and external information together.

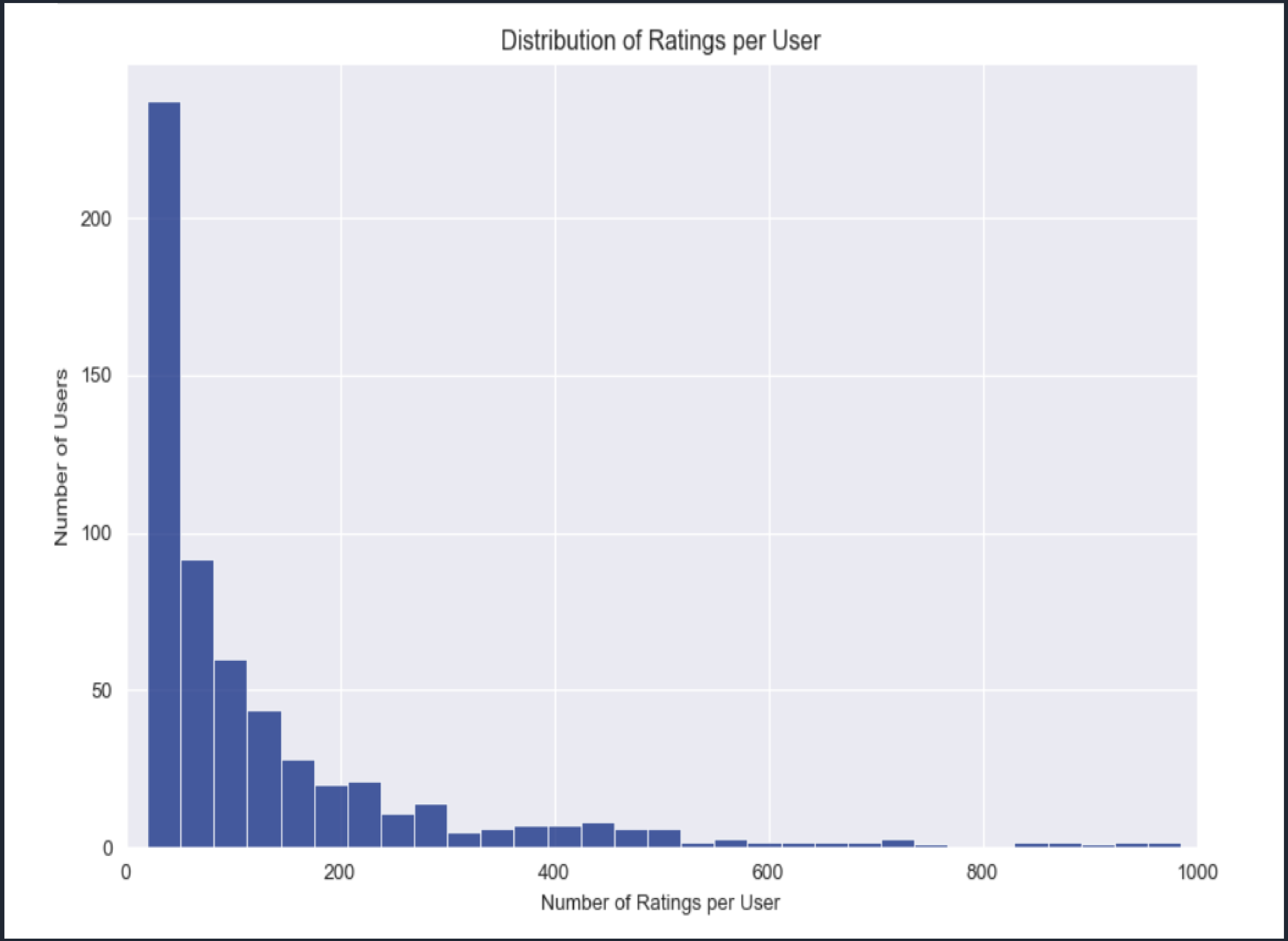
VISUALIZATIONS



Observation: Users tend to give positive ratings (many 4 and 5 stars) and use half-star ratings.

Insight for Modeling: This shows us we have good positive feedback to learn from and that users provide detailed preferences. It also suggests we need models that can handle variations in how users rate.

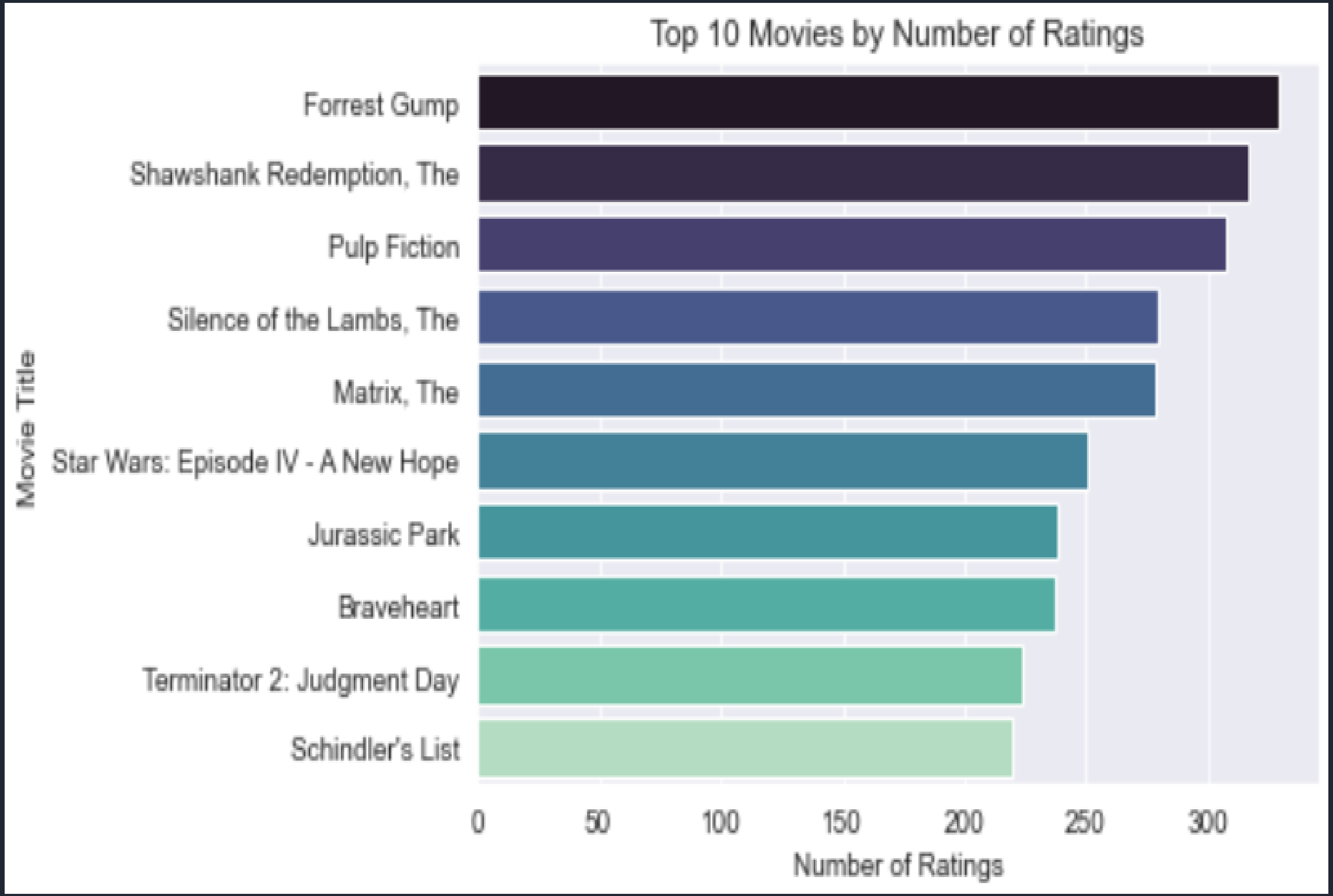
VISUALIZATIONS



Observation: Most users have only rated a small number of movies, while a few users rate very many.

Insight for Modeling: This clearly shows the "cold-start user" challenge – many users don't have enough history for standard recommendations. This insight was key to deciding we needed a hybrid system that can work even with limited user data.

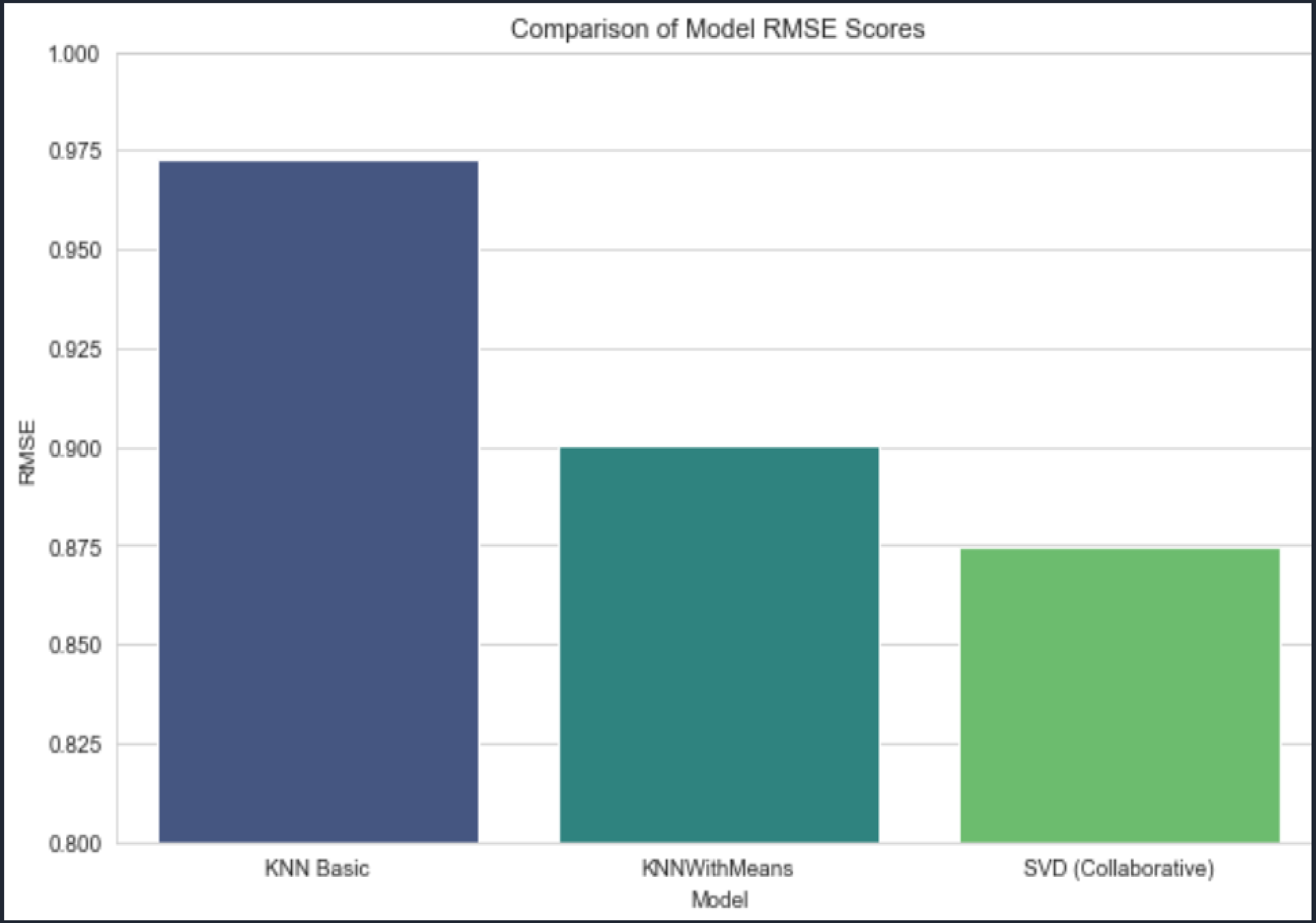
VISUALIZATIONS



Observation: A few movies receive a much larger number of ratings than others.

Insight for Modeling: These popular movies are important because they provide a lot of data for our main models to learn from. Understanding them also helps us think about how to recommend *new* movies that don't have many ratings yet (cold-start items).

RESULTS



The RMSE score is used to measure how close our predicted rating for a movie was to the actual rating a user gave. The lower the bar the better the accuracy. The goal is to get this number as low as possible.

Observation: The bar for the SVD model is the lowest.

Insight for Modeling: The SVD method was the most accurate in predicting user ratings among those tested (lower bar = better accuracy). This means our best model is effective at predicting what users will like, which helps us make good recommendations.

BUSINESS RECOMMENDATION

1

Cold-Start Strategy for New Users

Use content-based filtering to recommend movies to new users based on selected genres or movie preferences during onboarding. This mitigates cold-start issues and improves first-time user satisfaction.

2

Hybrid Approach for Balanced Recommendations

Implement a weighted or switching hybrid model to combine the strengths of collaborative filtering and content-based filtering, improving recommendation diversity and accuracy.

3

Promote Hidden Gems to Maximize Catalog Value

Leverage the model insights to surface highly rated but under-watched movies and push them to relevant users. This increases exposure for lesser-known content and improves catalog utilization across the platform.

4

Encourage High-Quality Tags

Encourage users to contribute more high-quality tags during their interaction with the platform (e.g., after rating a movie) which will help reduce noise and improve content discoverability especially for niche titles.

FUTURE WORKS

Leverage Implicit Feedback

Beyond explicit ratings, explore incorporating implicit user signals such as searching for specific genres, rewatching the same movie. It will provide a more complete picture of user behavior and can significantly enhance recommendation accuracy for users.

Investigate Dynamic Shifts

Investigate how user preferences and item popularity evolve over time. Implement time-aware techniques to capture these dynamic shifts in recommendations. Accounting for time ensures recommendations remain fresh and relevant, adapting to changing user preferences and new content releases.

Explore Advanced Modeling Techniques

Explore Advanced Modeling Techniques: such as Neural Collaborative Filtering or sequential models, to capture more complex patterns in user-item interactions. These advanced methods can reveal deeper insights into user preferences.

Expand Movie Features

Expand the set of features used to describe movies beyond genres and tags, potentially incorporating information from synopses or cast/crew data. More comprehensive content descriptions improve the system's ability to handle new or niche items.

CONCLUSION

This project demonstrates how a hybrid recommender system can address the content discovery problem on a platform like StreamTimeNow. By combining collaborative filtering and content-based techniques, we’re able to provide high-quality, personalized movie recommendations even for new users.

CONTACTS

For any additional questions, please contact these members:

- Calistus Mwonga, calistusmwonga@gmail.com
- Samwel Kipkemboi, samkemboi201@gmail.com
- Brian Kanyenje, bkanyenje@gmail.com
- Hannah Nyambura, anngachuhipg1@gmail.com
- Kelvin Mutua, kelvinmutua787@gmail.com