BUILDING A MOVIE RECOMMENDER SYSTEM

Group Members

- Esther Francis - Team Leader
- Doreen Wanjiru
- Gregory Mikuro
- Ian Korir

Overview

- Background
- Problem Statement
- Objectives
- Data Understanding and Preparation
- Exploratory Data Analysis
- Modeling
- Deployment
- Conclusion

Background

- In today's streaming era, with a staggering 8 billion hours of content consumed daily on Netflix alone, finding the perfect movie is a daunting task.
- Individual preferences for genres, actors, directors, and themes vary widely, making a one-size-fits-all recommendation approach ineffective.
- Streaming platforms collect vast amounts of data on user behavior (watch history, ratings, searches), which can be leveraged to understand and predict user preferences.
- This presentation describes a movie recommendation system creation process that enhances the user’s movie-watching experience by suggesting films tailored to their preferences.

Problem Statement

Users are overwhelmed by the sheer volume of available movies and struggle to sift through the options to find something they'll enjoy. Generic recommendations based on broad trends or popularity often result in users being presented with movies that don't align with their personal taste. Users spend valuable time browsing through endless lists of movies, leading to frustration and potentially abandoning the platform altogether. Without personalized guidance, users may miss out on discovering hidden gems that perfectly match their preferences.

Aim and Objectives

Primary Aim: To develop and implement a Movie Recommender System that delivers personalized and relevant movie suggestions to users, thereby improving their movie discovery experience and increasing their satisfaction with their entertainment choices.

Objectives:

1. Analyze and determine movie ratings considering the number of users who rated them.
2. Investigate relationships between user preferences and movie features through matrix factorization techniques.
3. Create a hybrid recommendation model that integrates collaborative filtering (with a target accuracy of 80% or higher) with content-based filtering.
4. Deploy the developed recommendation system using Streamlit.

Data Understanding and Preparation

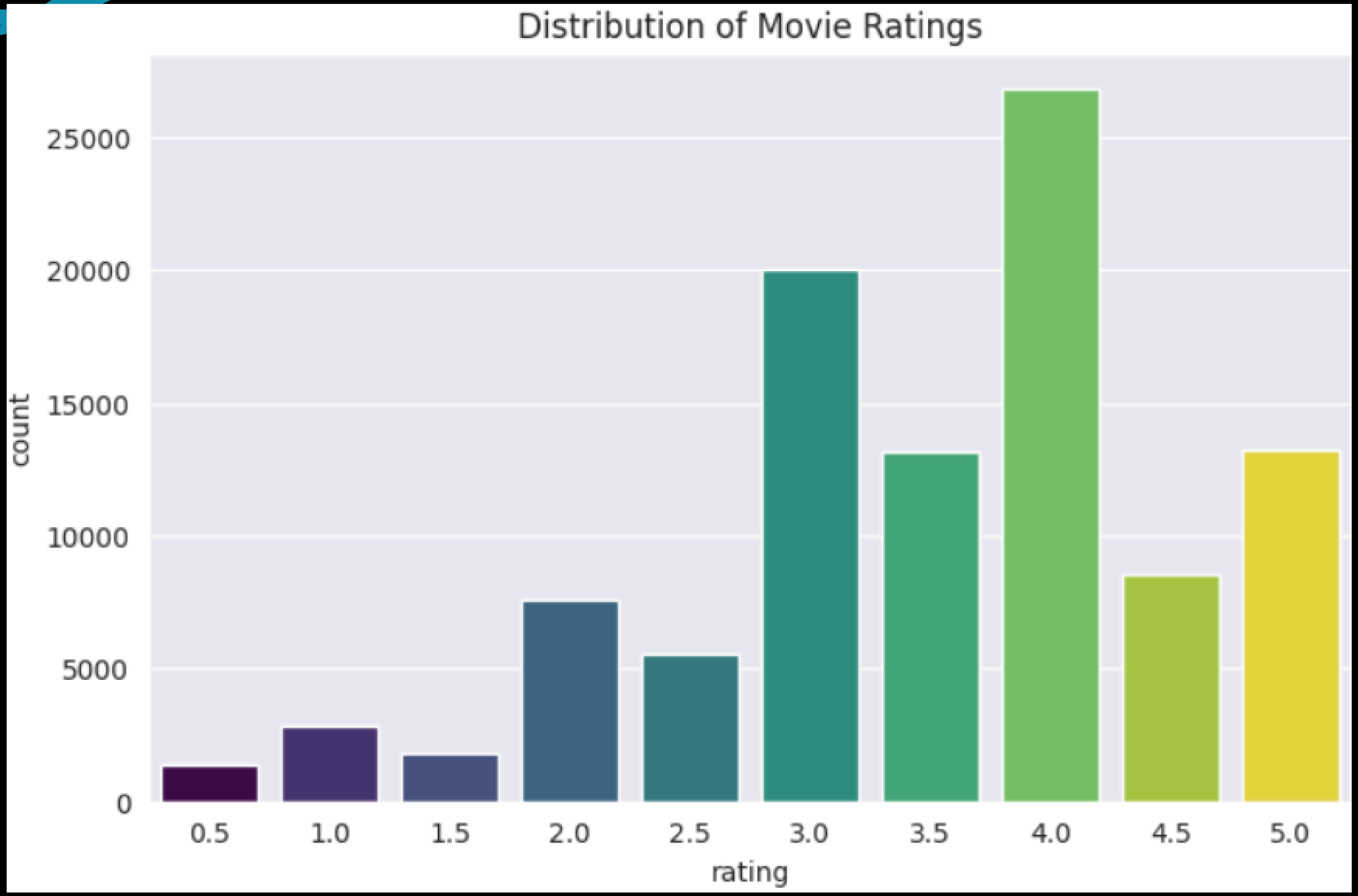
- The dataset ml-latest-small is from Here . which describes 5-star ratings and free-text tagging activity from movieLens, a movie recommendation system.
- Data contained 100836 ratings across 9742 movies and was generated on September 26th, 2018.
- Users were selected randomly for inclusion and all the users had rated at least 20 movies.

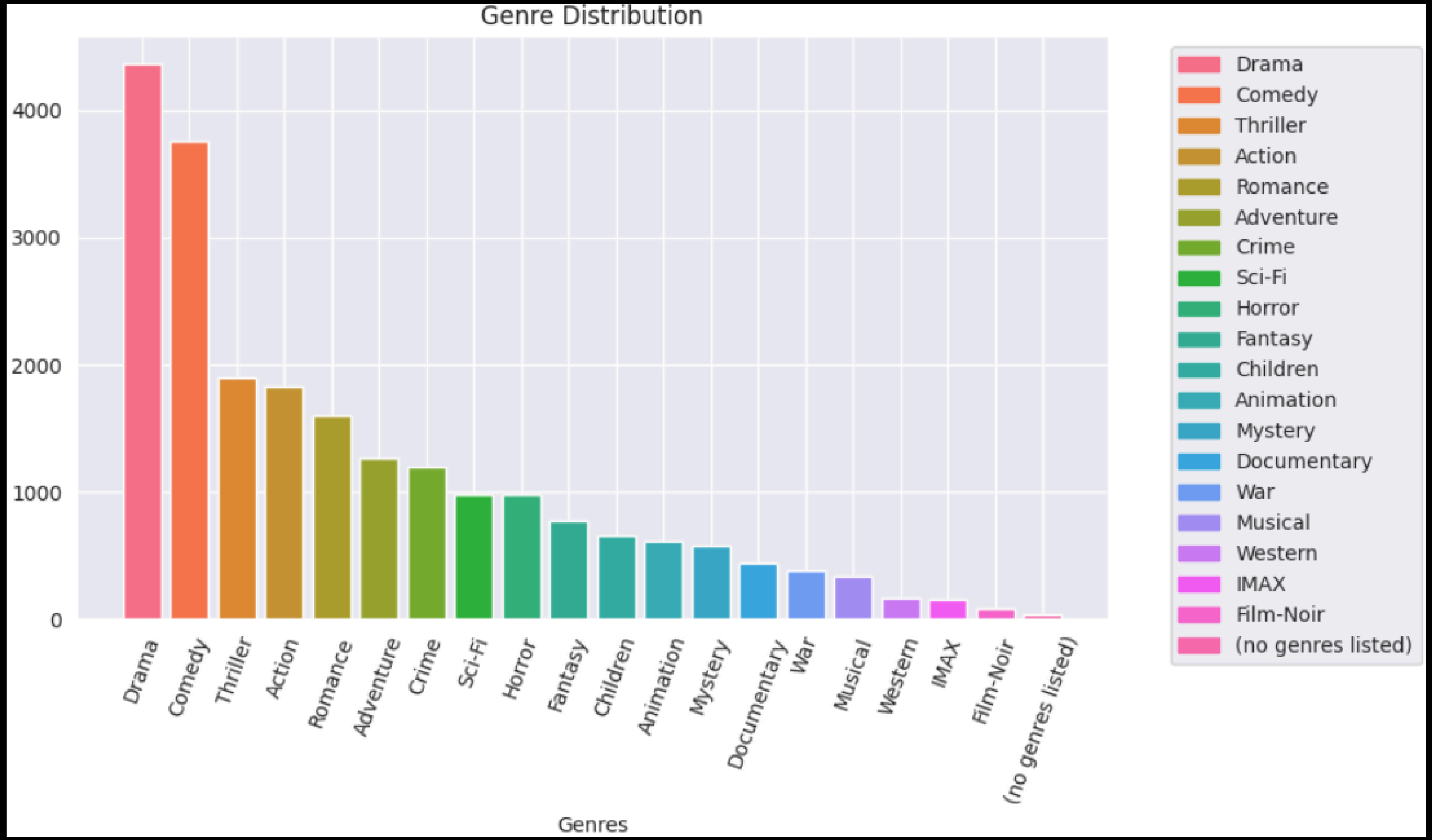
Data Preparation and Cleaning

- **Consolidate:** Merge ratings and movie data for comprehensive analysis.
- **Simplify:** Remove irrelevant timestamps to reduce data complexity.
- **Deduplicate:** Eliminate duplicate entries for unbiased analysis.
- **Transform Genres:** Split combined genres into individual lists for granular analysis.
- **Extract Year:** Separate release year from title for temporal insights.
- **Calculate Bayesian Average:** Use Bayesian Average for fairer movie rating comparisons.

Exploratory Data Analysis

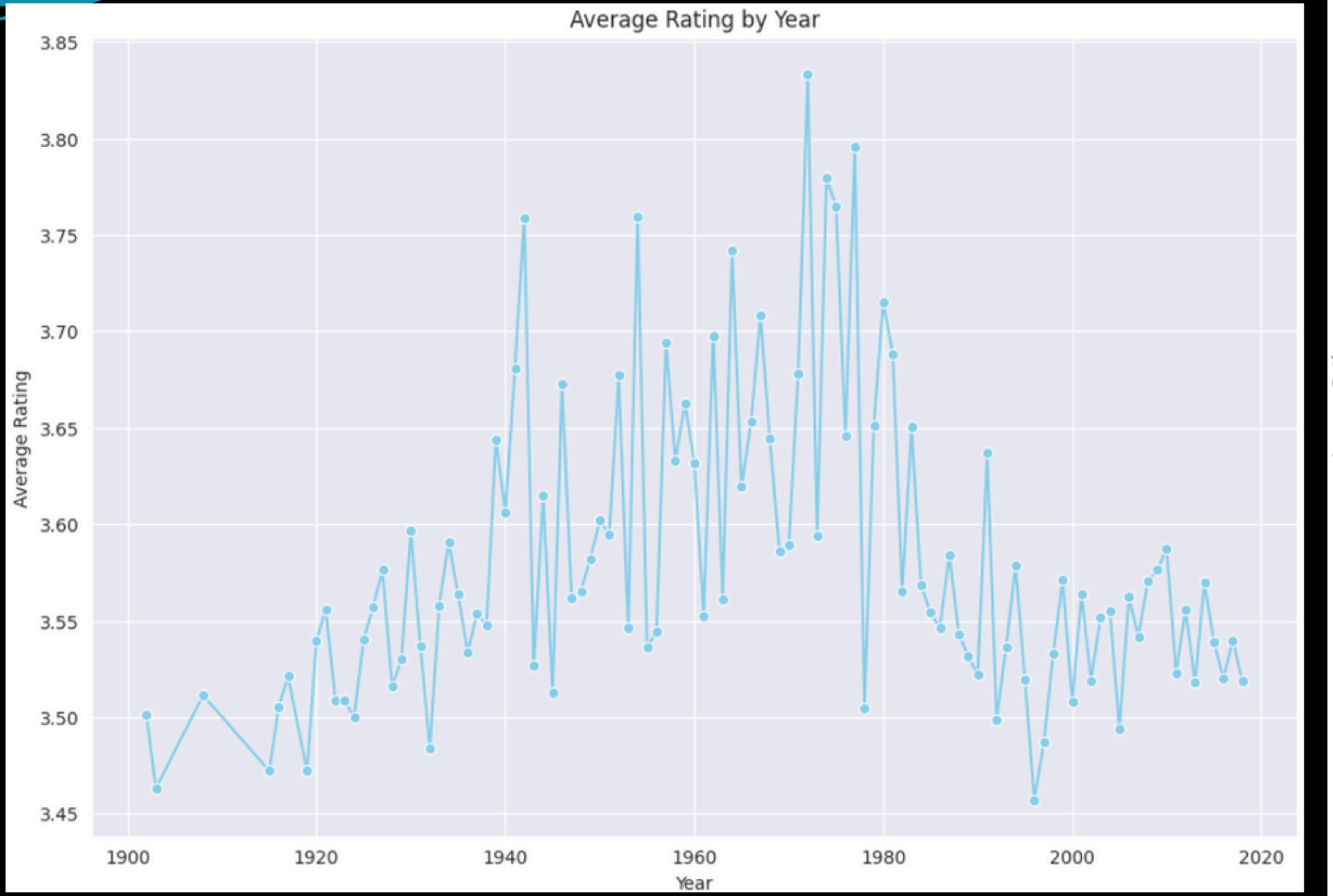
Univariate Analysis

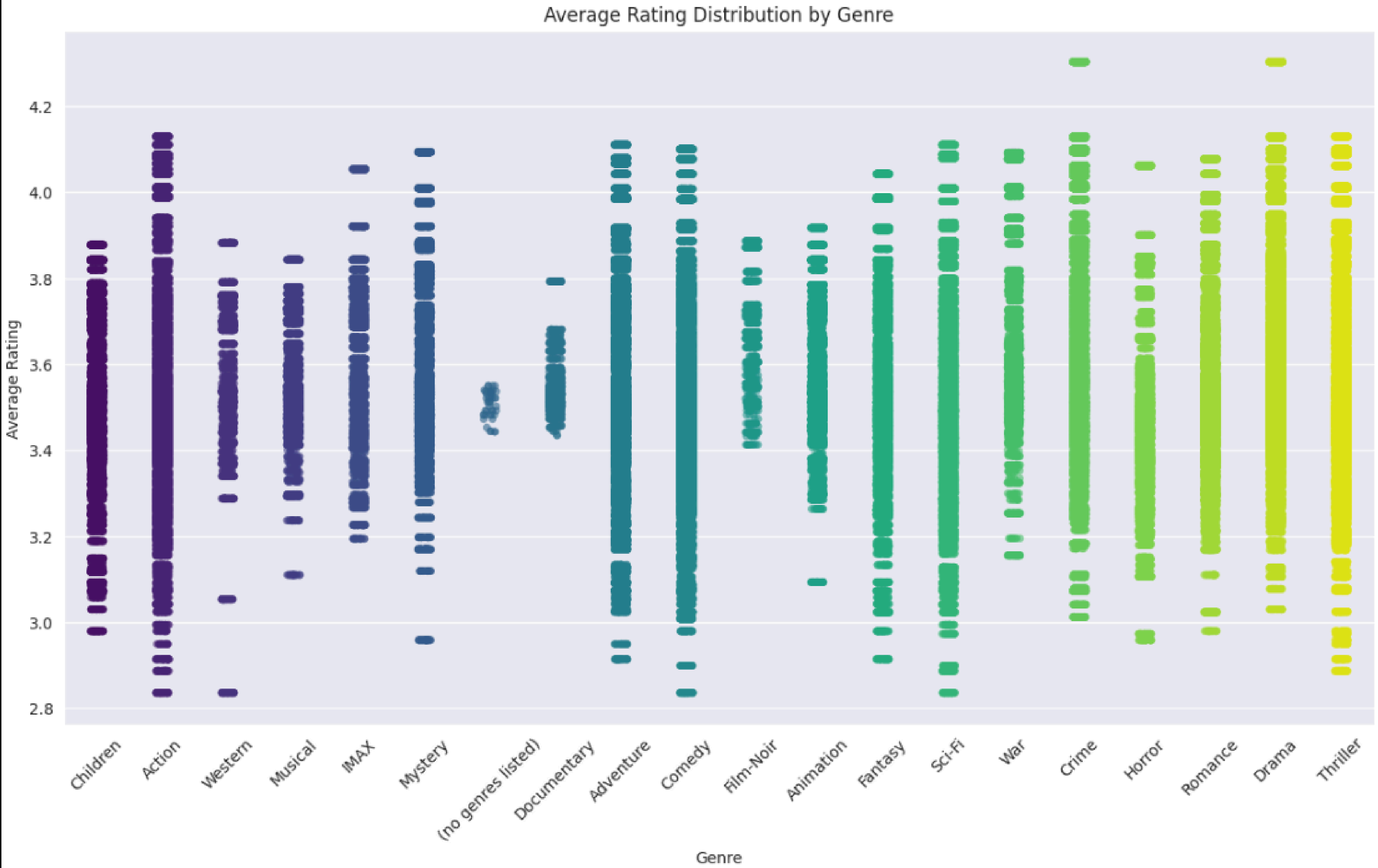




Exploratory Data Analysis

Bivariate Analysis

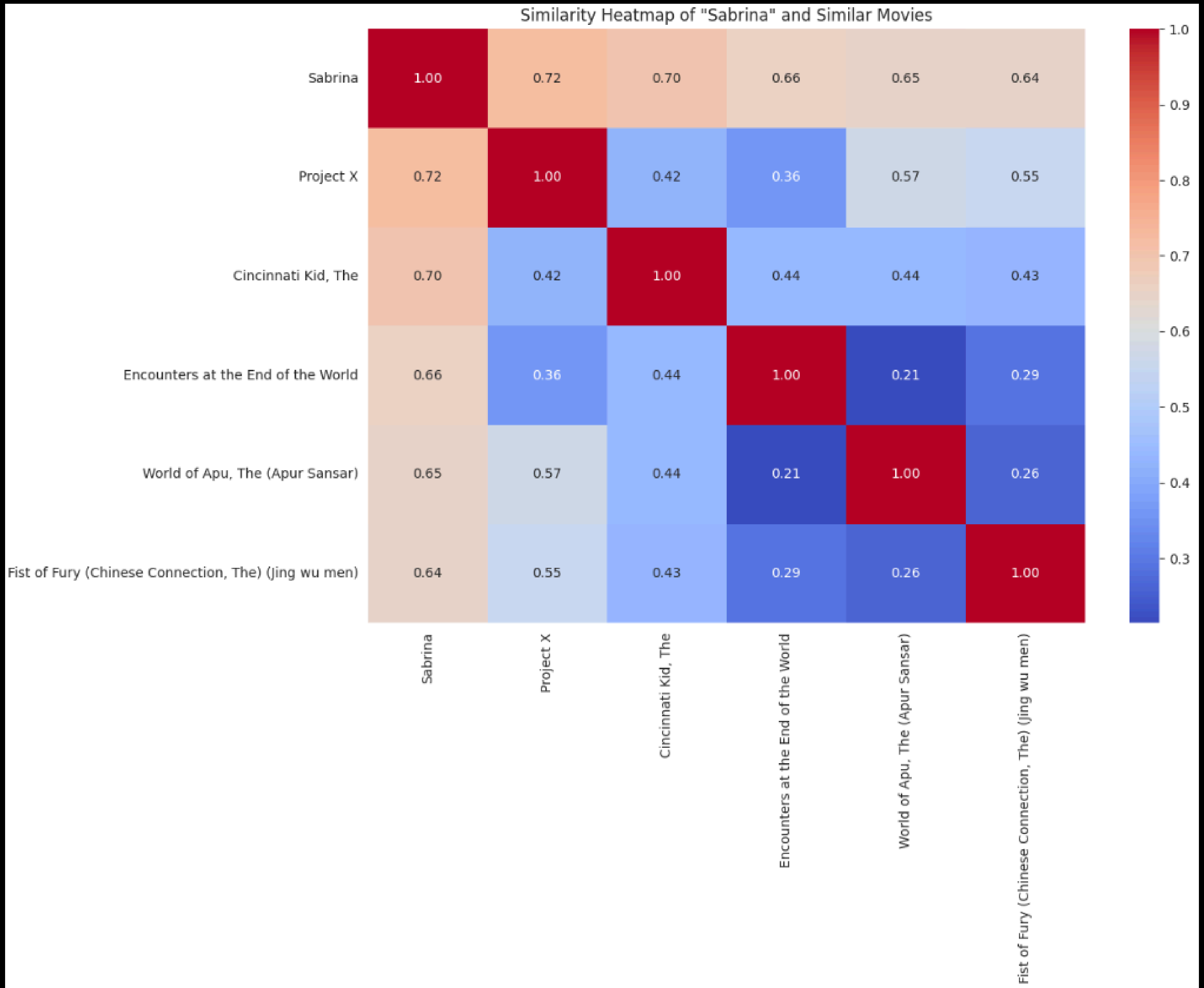




Exploratory Data Analysis

Multivariate Analysis

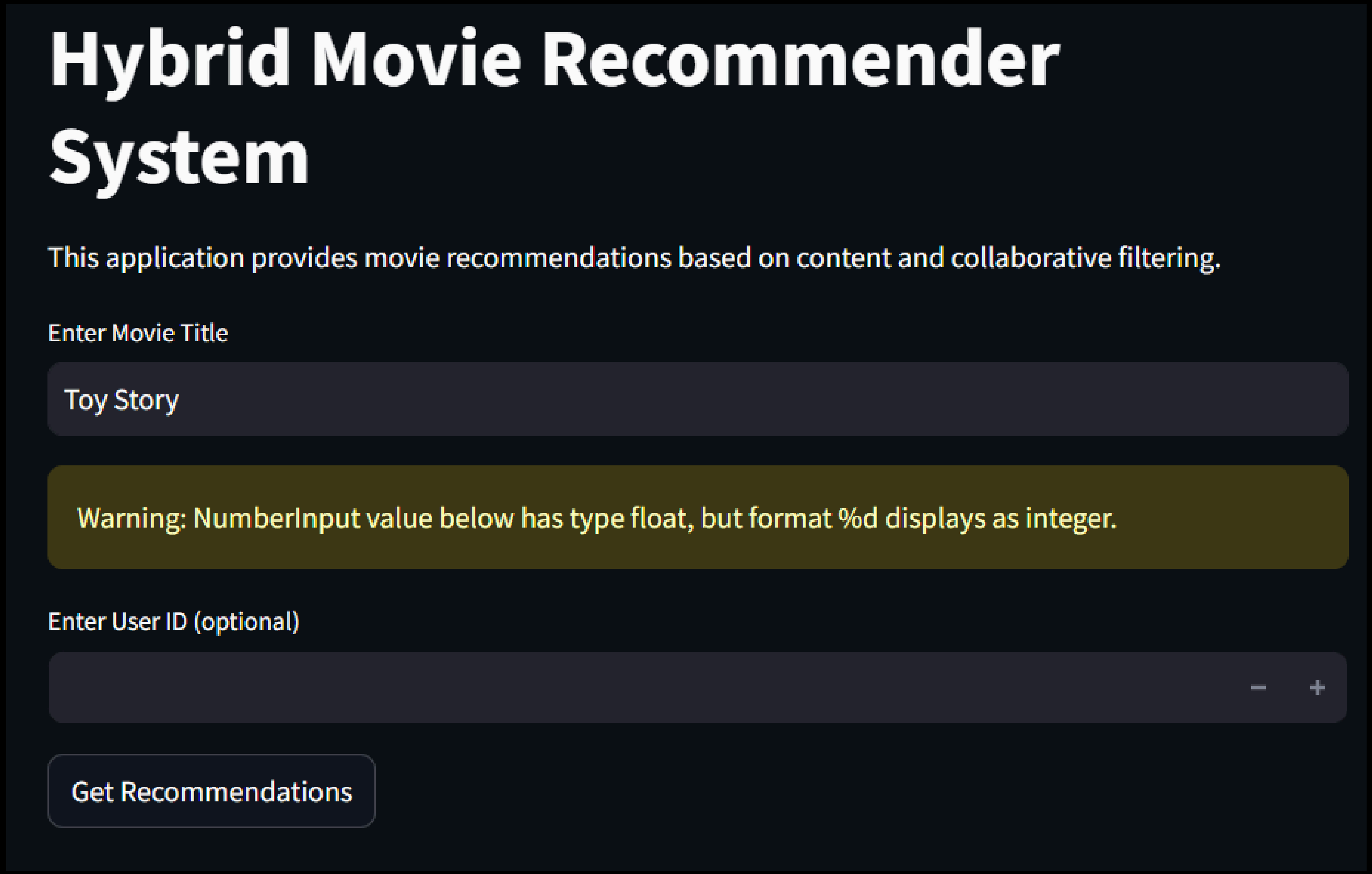
- The multivariate analysis encompassed processing movie data, training an SVD model for dimensionality reduction, and calculating cosine similarity between movie vectors.
- It provides methods to find the top similar movies to a given movie and visualize their similarity relationships using a heatmap.



Modeling

- A collaborative filtering model was developed to recommend movies based on user-item interactions and similarity metrics.
- A hybrid model was also implemented, combining collaborative filtering with content-based filtering using genre similarity.
- This hybrid approach selects the most suitable recommendation method based on user or movie ratings.
- Both models leverage data structures and algorithms for efficient processing of movie information, ultimately delivering personalized recommendations to users.

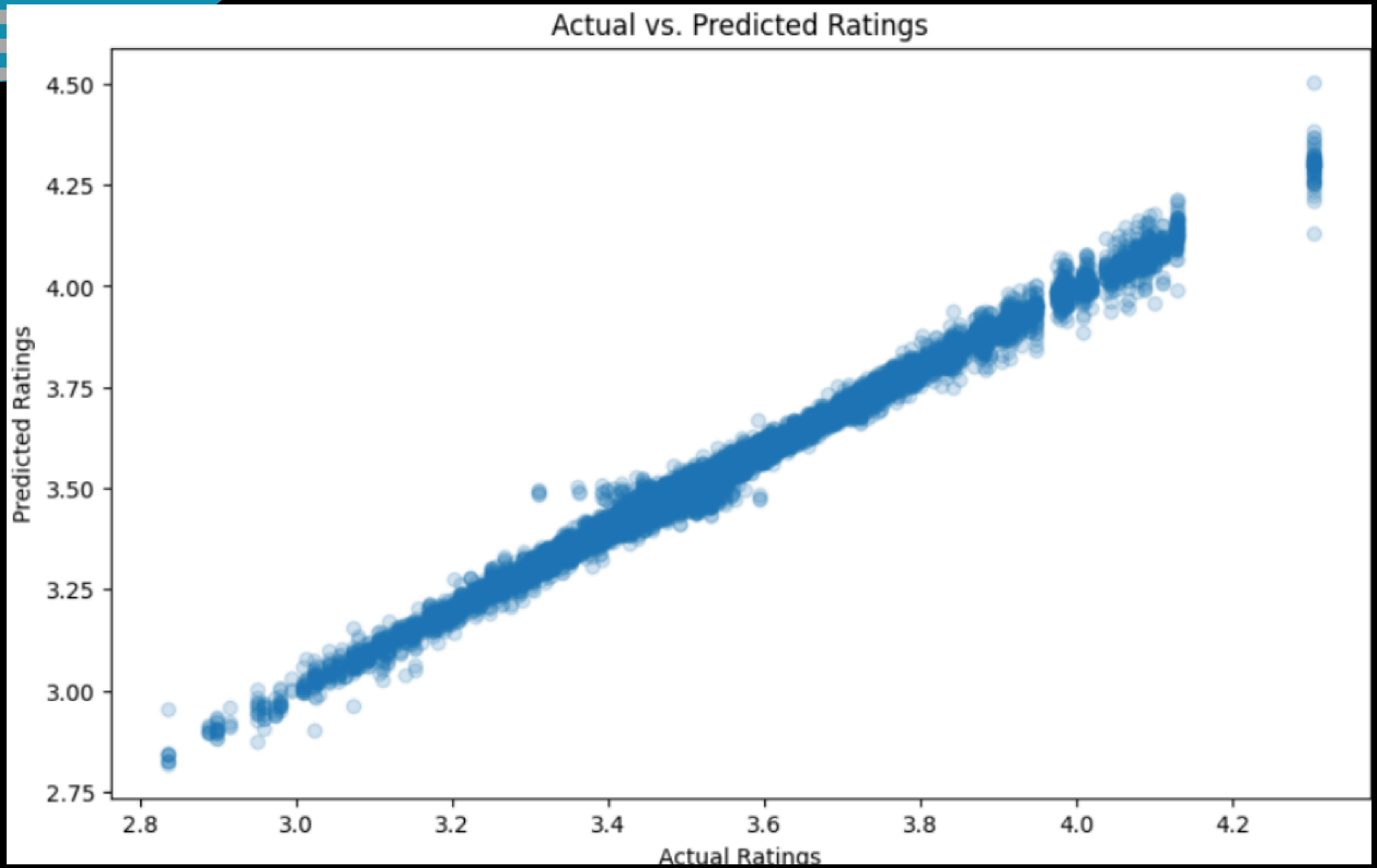
Deployment

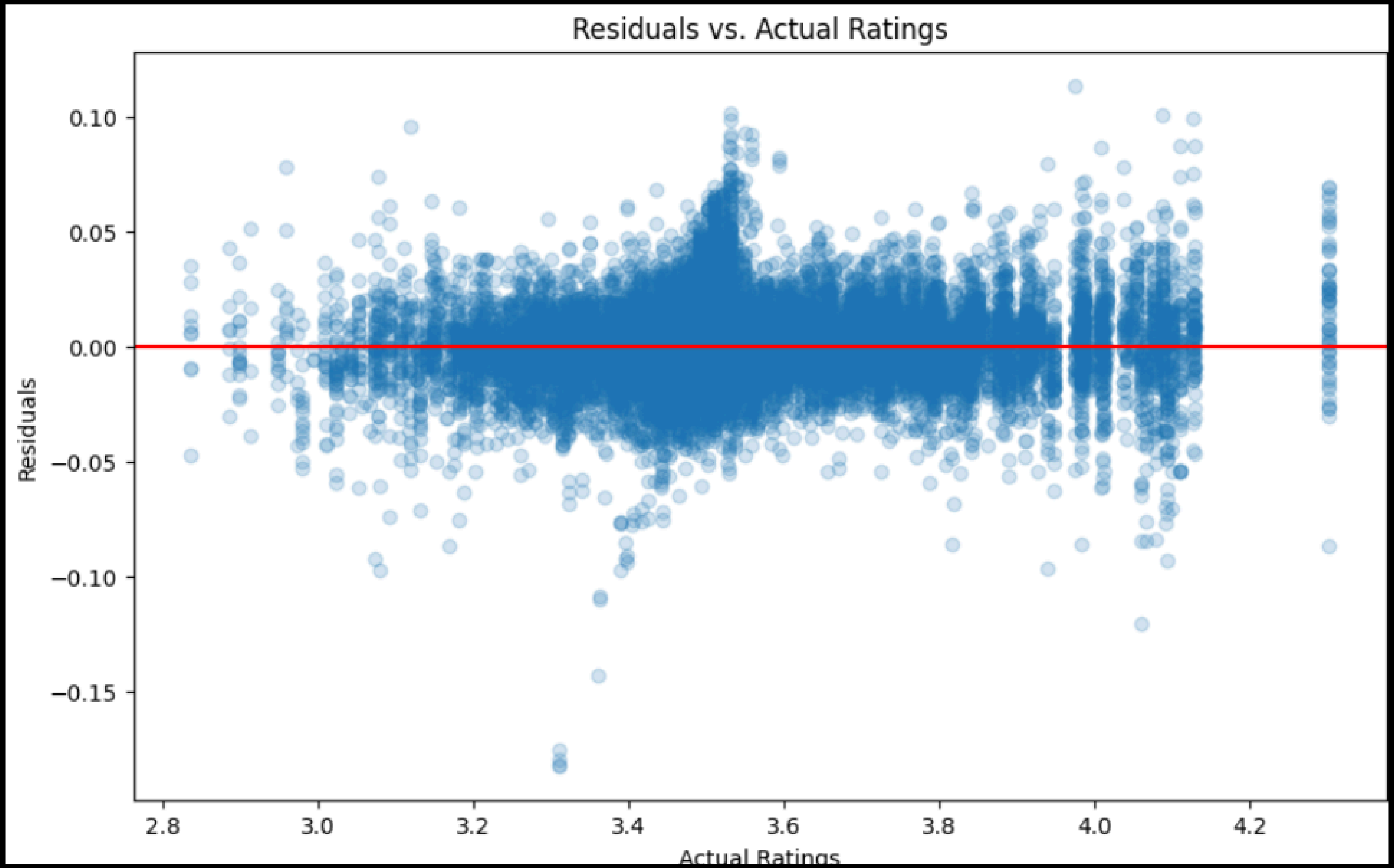


SVD-Based Recommender System - Calculates cosine similarity to identify and visualize similar movies

Hybrid Recommender System - Decides whether to use content-based or collaborative filtering based on the user's rating history or the movie's rating history.

Evaluation


Actual vs. Predicted Ratings
Predicted Ratings
Actual Ratings
4.50
4.25
4.00
3.75
3.50
3.25
3.00
2.75
2.8
3.0
3.2
3.4
3.6
3.8
4.0
4.2


Residuals vs. Actual Ratings
Residuals
Actual Ratings
0.10
0.05
0.00
−0.05
−0.10
−0.15
2.8
3.0
3.2
3.4
3.6
3.8
4.0
4.2

Conclusion

- This project successfully developed a hybrid movie recommendation system that addresses the "choice overload" problem users face.
- The system leverages Bayesian averages for reliable movie ratings, matrix factorization to uncover user preferences, and a deep learning-based collaborative filtering model for personalized recommendations.
- Content-based filtering is incorporated to handle new users, and the system is deployed via an interactive Streamlit interface.
- Future enhancements include continuous model updates, expanding content features, gathering user feedback, and exploring contextual recommendations.

Thank You