BUILDING A MOVIE RECOMMENDER SYSTEM

Overview

- Problem Statement
- Business Understanding
- General Objectives
- Data Understanding
- Data Preparation
- Data Analysis and Visualization
- Modeling
- Deployment
- Conclusion

Problem Statement

In the vast and ever-expanding landscape of film and television content, users often face the challenge of discovering movies that align with their personal tastes and preferences. With countless options available across multiple platforms, finding enjoyable and relevant movies can be overwhelming and time-consuming. Traditional methods of browsing and searching are often inefficient, leading to decision fatigue and a suboptimal viewing experience. We aim to develop a Movie Recommender System that addresses this problem by providing users with personalized, relevant, and timely movie recommendations.

Business Understanding

In the age of streaming services and an abundance of movie alternatives, customers frequently face the issue of finding films that match their preferences. This project aims to build a movie recommendation system that enhances the user's movie-watching experience by suggesting films tailored to their preferences. The system will utilize machine learning algorithms and user data to deliver tailored movie suggestions based on ratings, viewing history, and preferences, resulting in growth and competitive advantage in the entertainment industry. The target audience for this project is companies that provide movie streaming services, such as Netflix, Amazon Prime Video, or Hulu. These companies can, in turn, use recommendation systems to increase their customer engagement and retention

General Objectives

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.

Data Understanding

The dataset ml-latest-small is from

(https://grouplens.org/datasets/movielens/latest/)

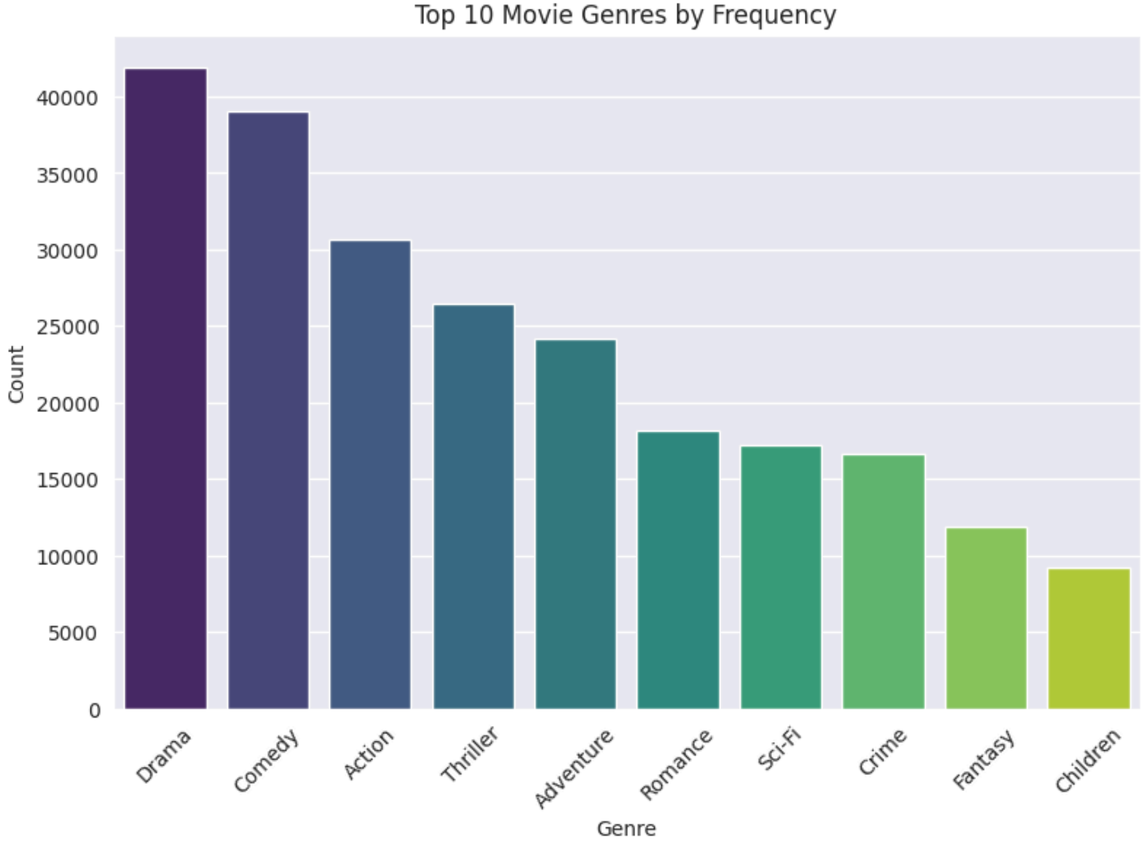
It describes 5-star ratings and free-text tagging activity from movieLens, a movie recommendation system. The data contained 100836 ratings across 9742 movies and was generated on September 26th, 2018. Users were selected randomly for inclusion and all the users had rated at least 20 movies.

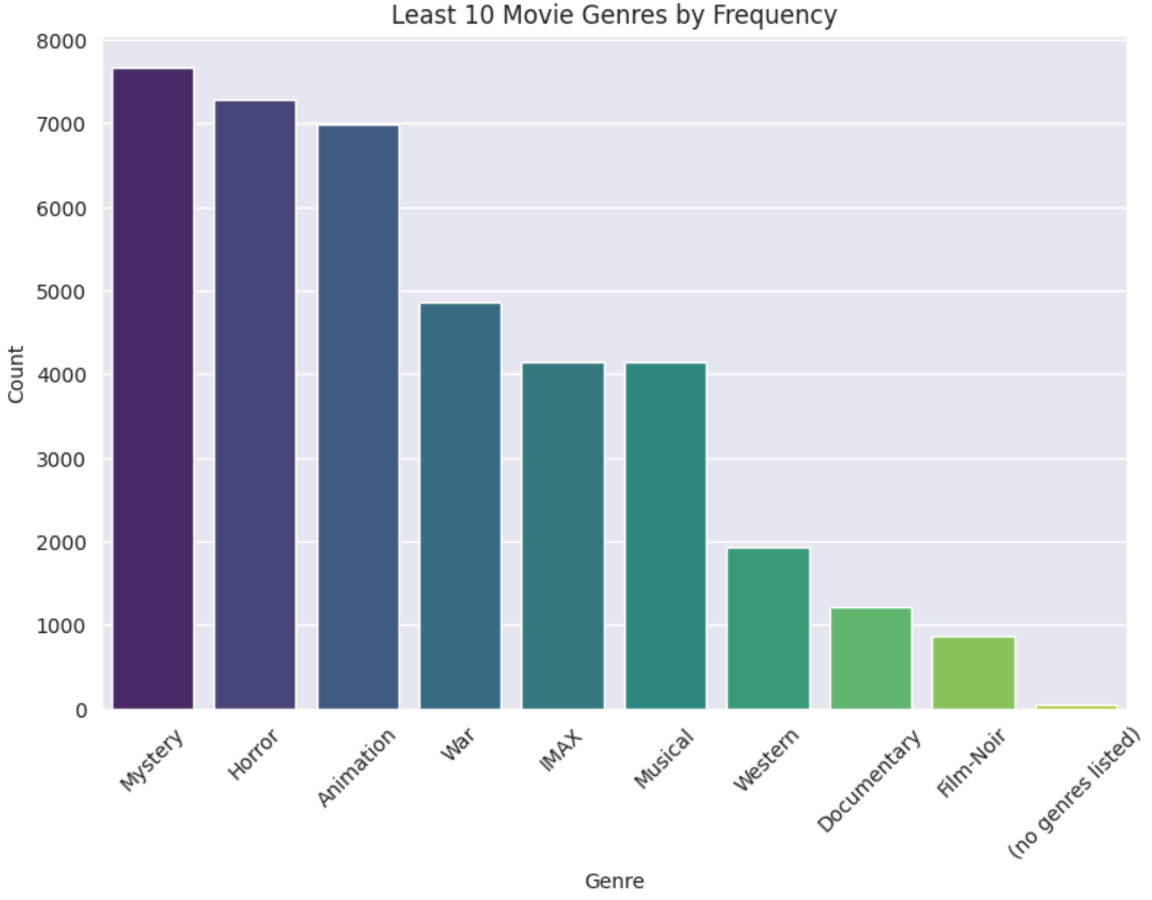
Data Preparation

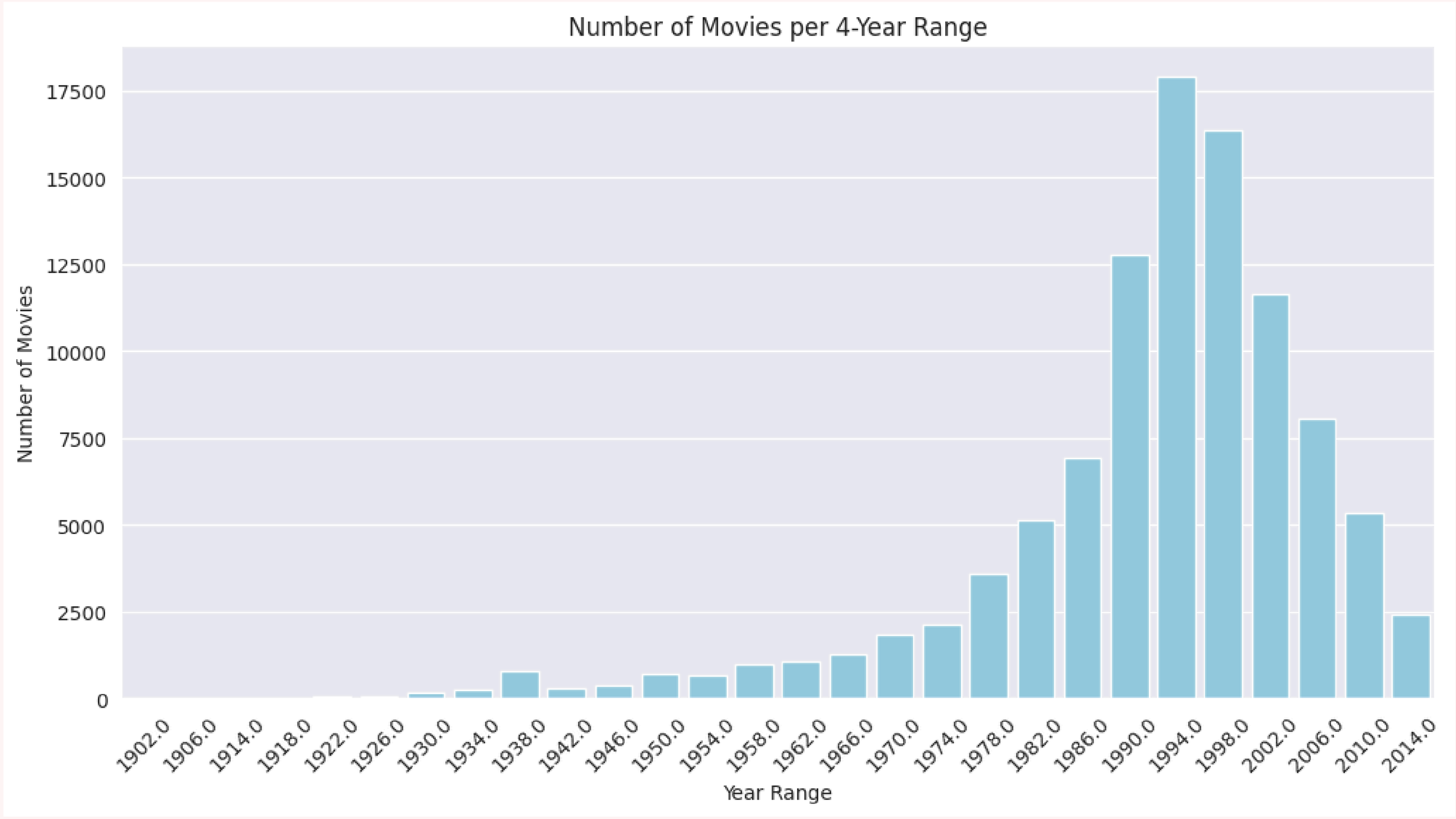
Summary of Features in the Dataset

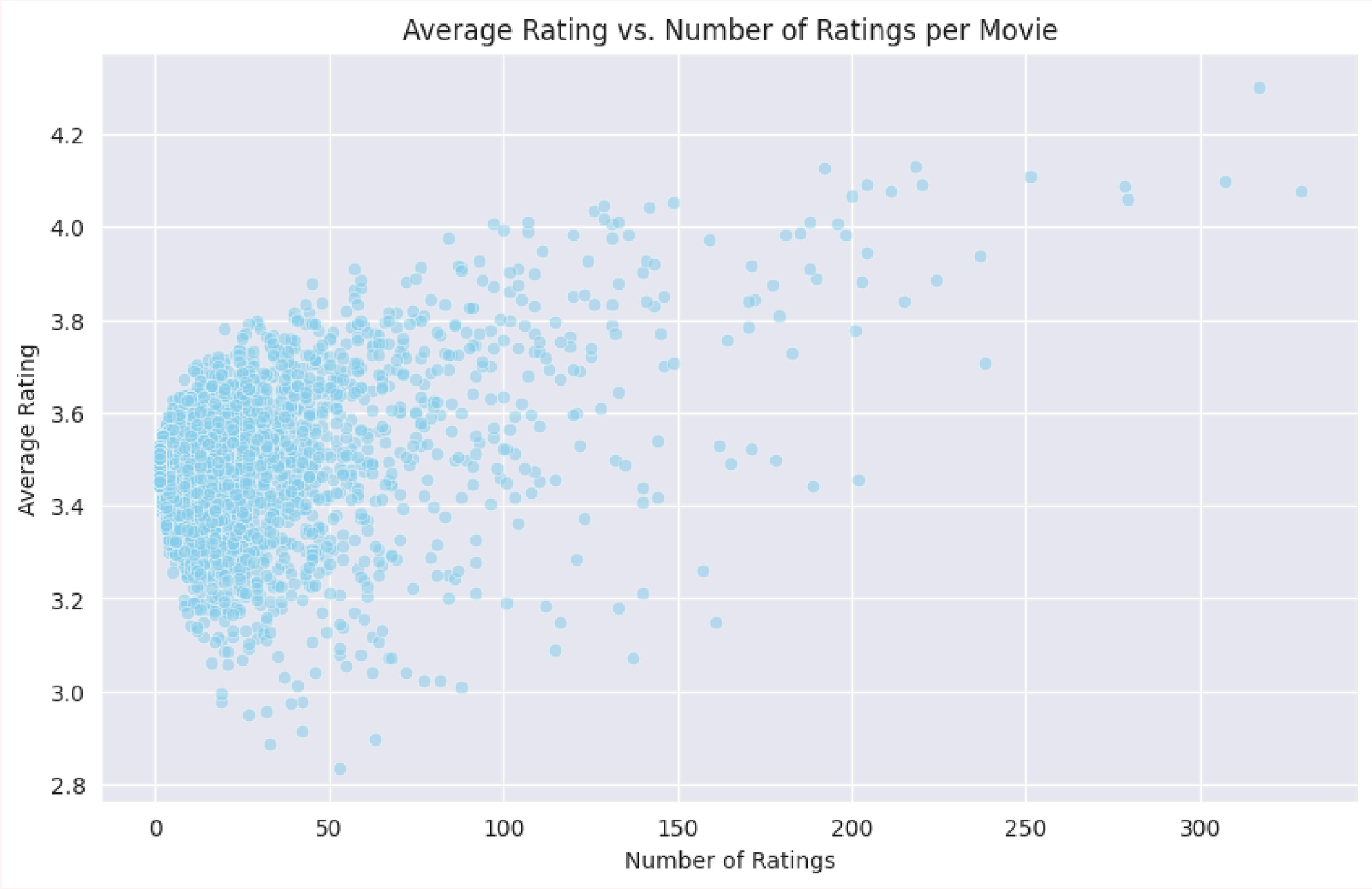
1. **User_Id:** A unique identifier for each user
2. **MovieId:** A unique identifier for each movie and is consistent in the dataset in ratings, tags, movies, and links.
3. **TimeStamp**: Represents seconds since midnight in Coordinated Universal Time (UTC).
4. **Tags:** User-generated metadata about movies. Each tag is typically a single word or short phrase where each user determines a particular tag's meaning, value and purpose.
5. **Genre:** Pipe-separated list, selected from Actions, Adventure, Animation, Children’s, Comedy, Documentary, Drama, Fantasy, Film-Noir, Horror, Musical, Mystery, Romance, Sci-Fic, Thriller, War, Western, no genres listed.

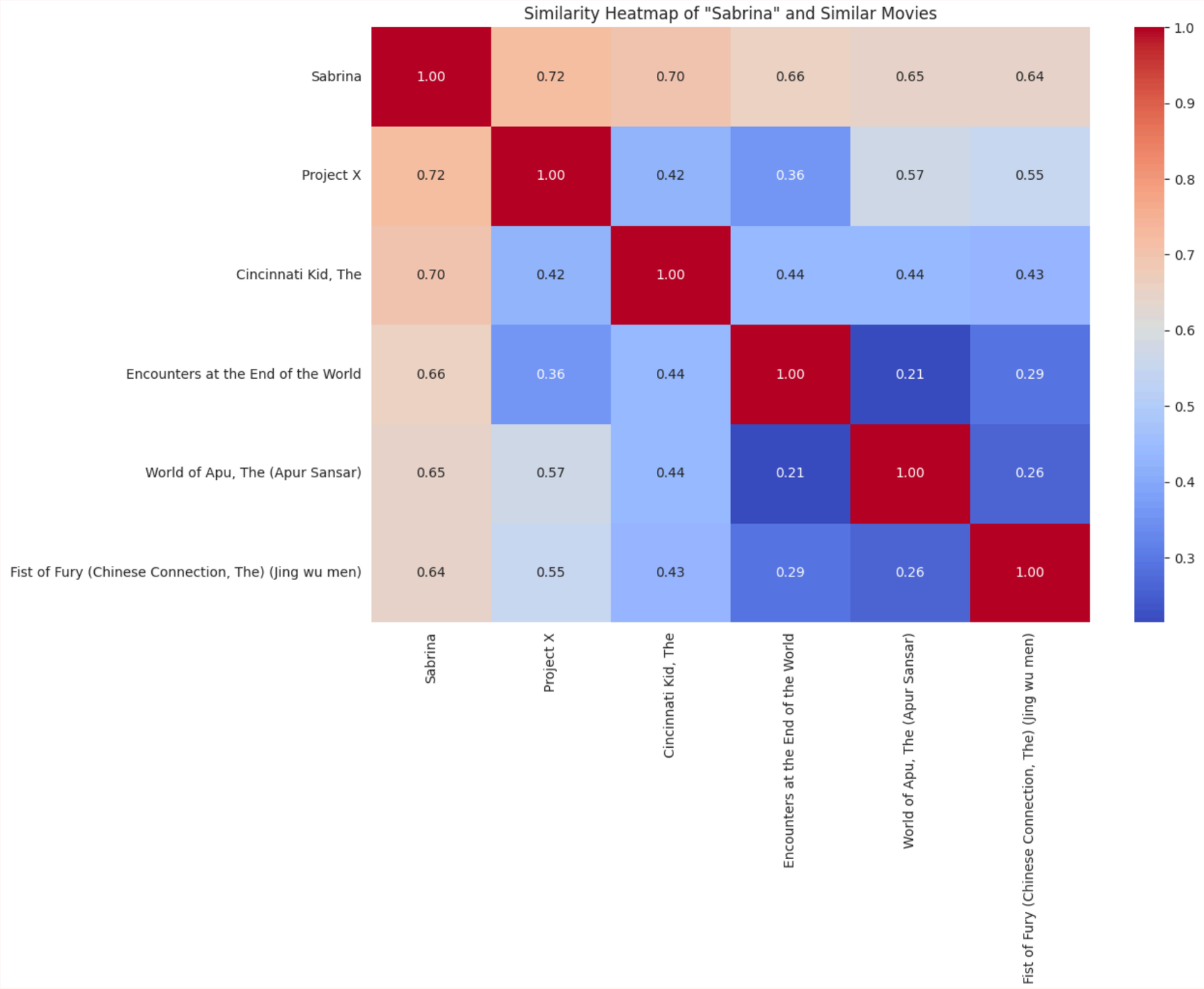
Data Analysis and Visualization


Top 10 Movie Genres by Frequency
Count
Genre
0
5000
10000
15000
20000
25000
30000
35000
40000
Drama
Comedy
Action
Thriller
Adventure
Romance
Sci-Fi
Crime
Fantasy
Children


Least 10 Movie Genres by Frequency
Count
Genre
0
1000
2000
3000
4000
5000
6000
7000
8000
Mystery
Horror
Animation
War
IMAX
Musical
Western
Documentary
Film-Noir
(no genres listed)


Number of Movies per 4-Year Range
Number of Movies
17500
15000
12500
10000
7500
5000
2500
0
1902.0
1906.0
1914.0
1918.0
1922.0
1926.0
1930.0
1934.0
1938.0
1942.0
1946.0
1950.0
1954.0
1958.0
1962.0
1966.0
1970.0
1974.0
1978.0
1982.0
1986.0
1990.0
1994.0
1998.0
2002.0
2006.0
2010.0
2014.0
Year Range


Average Rating vs. Number of Ratings per Movie
Average Rating
4.2
4.0
3.8
3.6
3.4
3.2
3.0
2.8
0
50
100
150
200
250
300
Number of Ratings


Similarity Heatmap of "Sabrina" and Similar Movies
Sabrina
Project X
Cincinnati Kid, The
Encounters at the End of the World
World of Apu, The (Apur Sansar)
Fist of Fury (Chinese Connection, The) (Jing wu men)
1.00 0.72 0.70 0.66 0.65 0.64
0.72 1.00 0.42 0.36 0.57 0.55
0.70 0.42 1.00 0.44 0.44 0.43
0.66 0.36 0.44 1.00 0.21 0.29
0.65 0.57 0.44 0.21 1.00 0.26
0.64 0.55 0.43 0.29 0.26 1.00
Sabrina
Project X
Cincinnati Kid, The
Encounters at the End of the World
World of Apu, The (Apur Sansar)
Fist of Fury (Chinese Connection, The) (Jing wu men)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3

Modeling

The hybrid_recommendations method decides whether to use content-based or collaborative filtering based on the user's rating history or the movie's rating history. It applies filtering based on thresholds for user and movie ratings to choose the most appropriate recommendation approach.

The hybrid recommendation system has successfully provided a diverse set of movie suggestions based on user history and content-based filtering. The recommendations, demonstrate the model's capability to blend collaborative and content-based approaches effectively. By integrating user preferences with movie attributes, the system enhances the relevance of suggested titles. This method not only addresses the cold start problem but also ensures that the recommendations cater to various user tastes

Deployment

- SVD App - The hybrid_recommendations method decides whether to use content-based or collaborative filtering based on the user's rating history or the movie's rating history. It applies filtering based on thresholds for user and movie ratings to choose the most appropriate recommendation approach.

- Hybrid Recommender App - The hybrid_recommendations method decides whether to use content-based or collaborative filtering based on the user's rating history or the movie's rating history. It applies filtering based on thresholds for user and movie ratings to choose the most appropriate recommendation approach.

Conclusion

In this project, we developed and evaluated several recommendation models based on the MovieLens dataset. We implemented collaborative filtering and hybrid recommendation systems to address various aspects of movie recommendation. The Singular Value Decomposition (SVD) model and the Hybrid Recommender were deployed using Streamlit, providing a user-friendly interface for generating and visualizing movie recommendations.

However, the evaluation of these models was not performed in detail due to limitations in computational power and potential biases introduced by the data splitting process. Specifically, the split did not fully capture the variability in user ratings or movie preferences, leading to skewed performance metrics. The true effectiveness of the recommendations could be more accurately assessed through continuous data collection and analysis of user feedback on the recommendations over time.

Any Questiions?