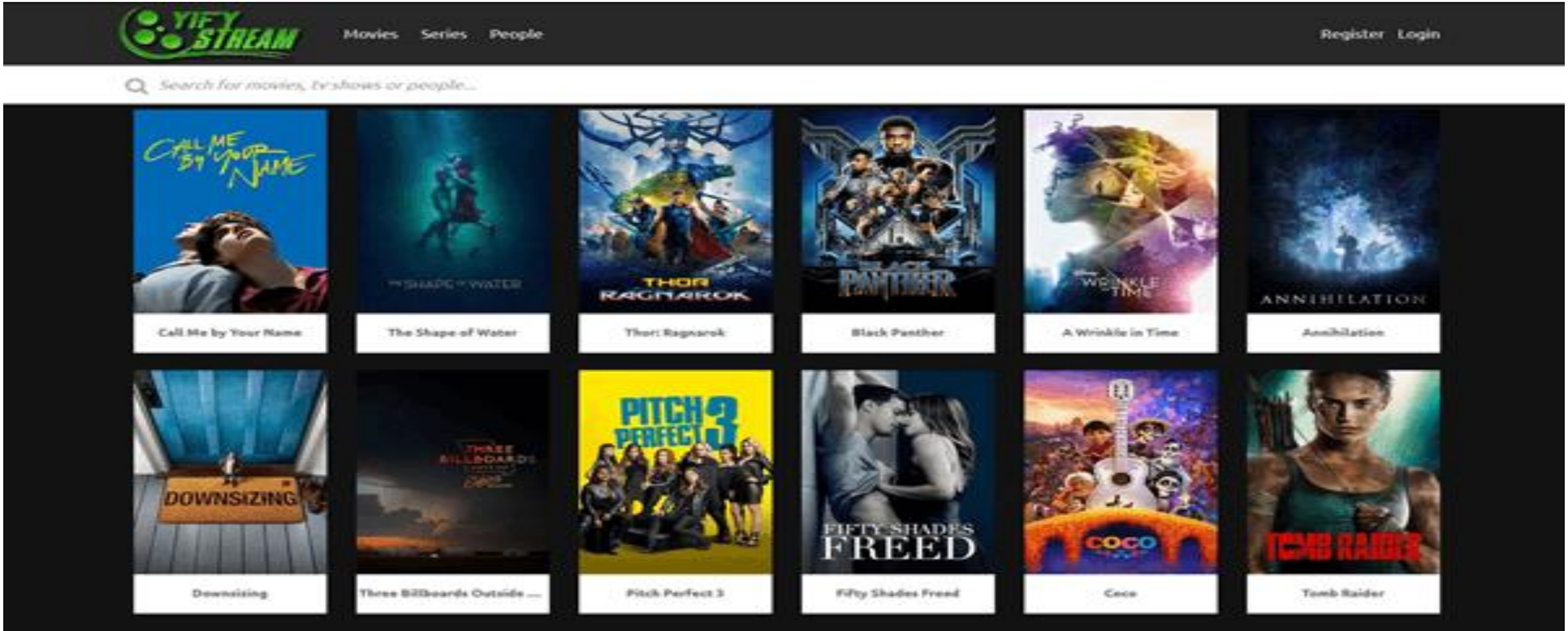
# MOVIES RECOMMENDATION SYSTEM



# PROBLEM STATEMENT




- ❖With vast movie catalogs, it can be overwhelming for users to find films they enjoy, leading to lower engagement and subscription cancellations.

- ❖An intelligent movie recommendation system can help users discover films that match their tastes, enhancing user satisfaction and platform engagement.

# STAKEHOLDERS

The primary stakeholders are:

- ❖**Streaming platforms** : They benefit from increased user engagement and retention by offering personalized recommendations.

- ❖**Users** : They receive tailored suggestions, improving their movie-watching experience.

# OBJECTIVES

**General Objective**

- ❖ Build a recommendation system that improves user engagement by providing relevant movie suggestions.

**Specific Objectives**

- ❖To build  a collaborative filtering model  that recommends movies to users based on their previous ratings and the behavior of similar users.
- ❖ To use past user ratings to predict ratings for unrated movies.
- ❖ To provide personalized movies recommendations.

# DATA UNDERSTANDING

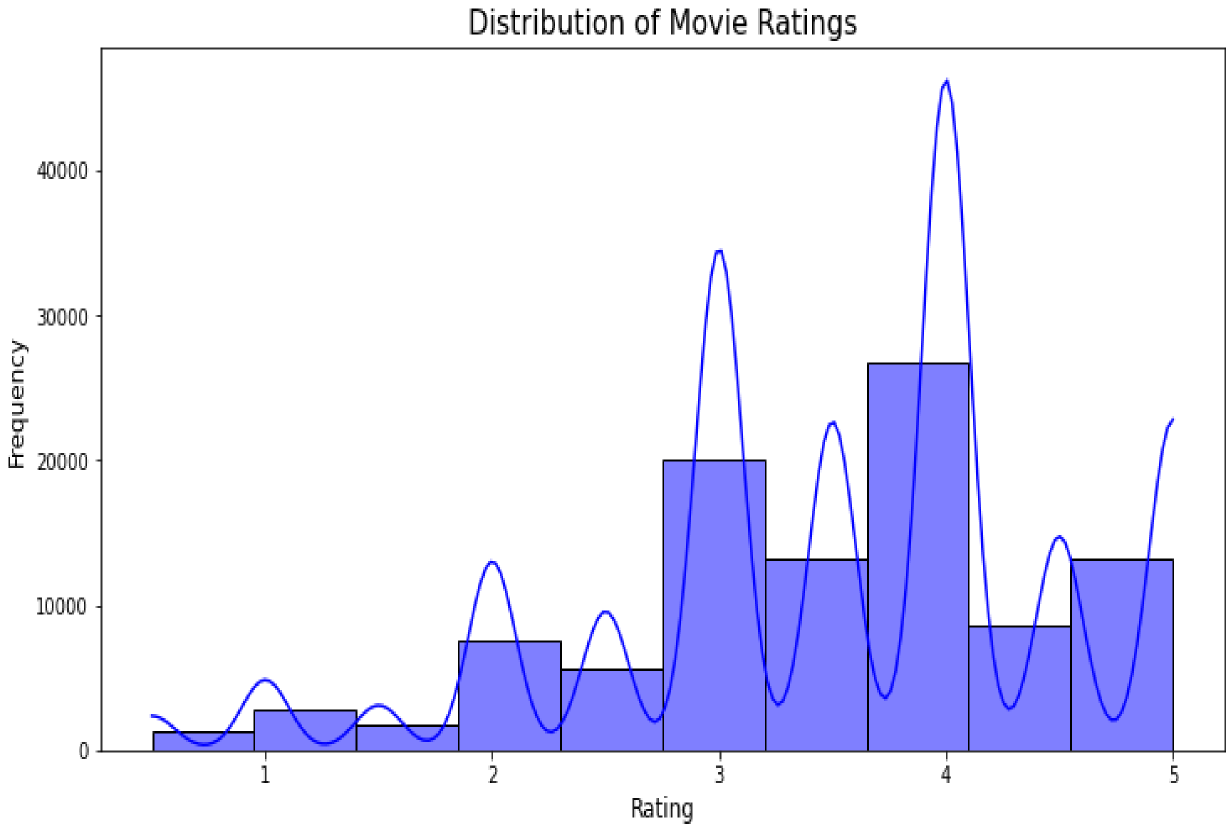
**Data Source**

The data for this project comes from MovieLens small dataset which is widely-used as a benchmark dataset provided by the GroupLens research group at the University of Minnesota. This dataset is designed specifically for building and testing recommendation systems, making it highly relevant for our project.

The dataset contains the following files:

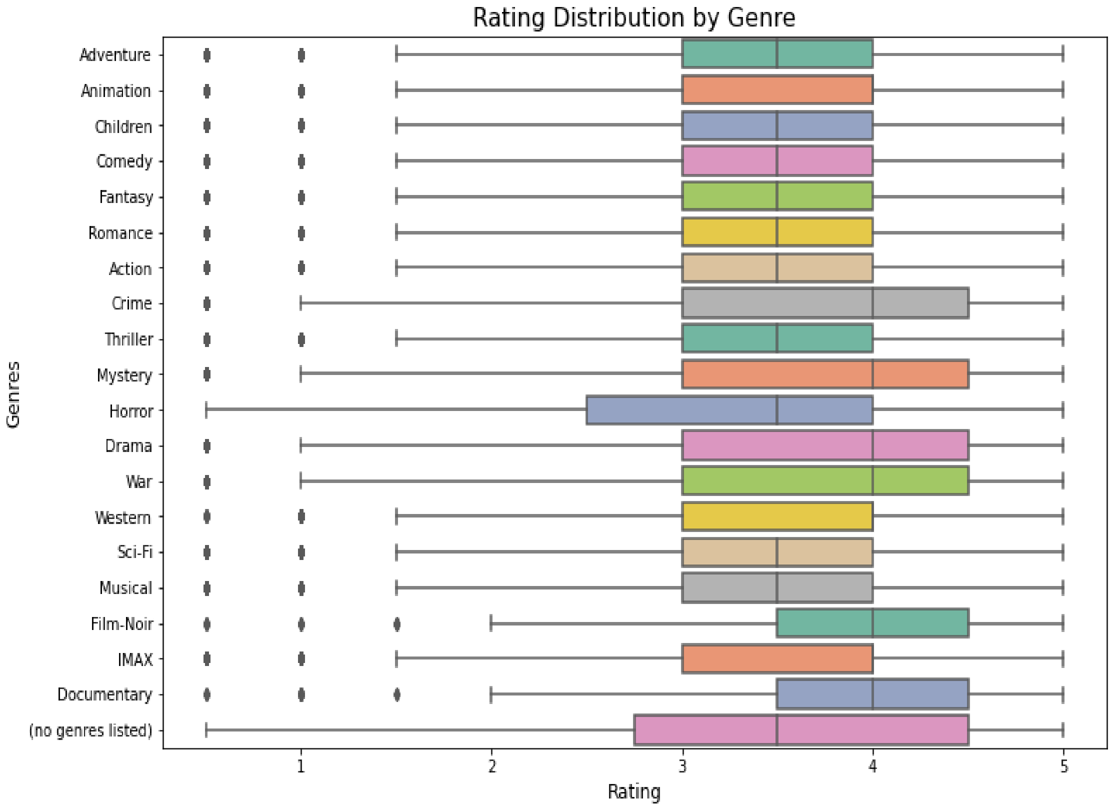
- ❖**Ratings**: User ratings for different movies.
- ❖**Movies**: Information about movie titles and genres.
- ❖**Tags**: User-generated tags for movies.
- ❖**Links**: Metadata connecting movies to external resources like IMDb.

# FINDINGS



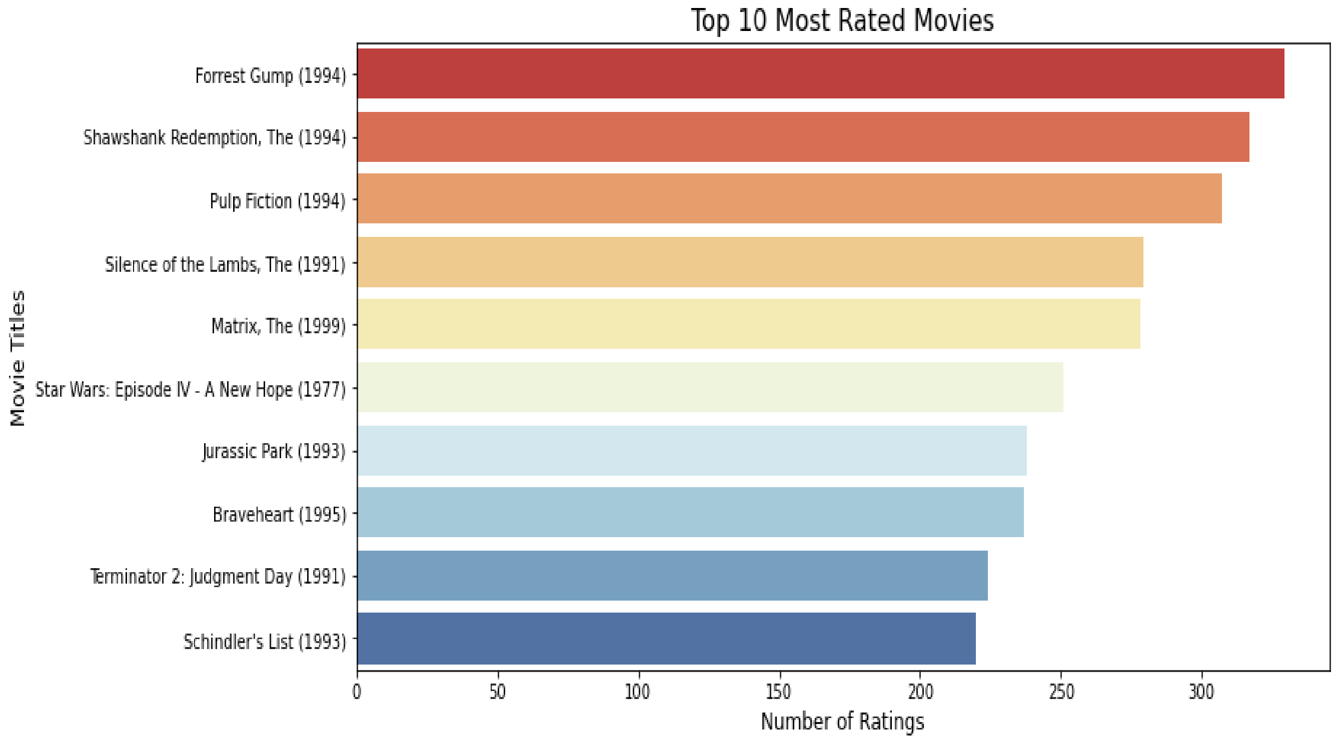

Most of the movies were rate between 3 and 4.

# FINDINGS



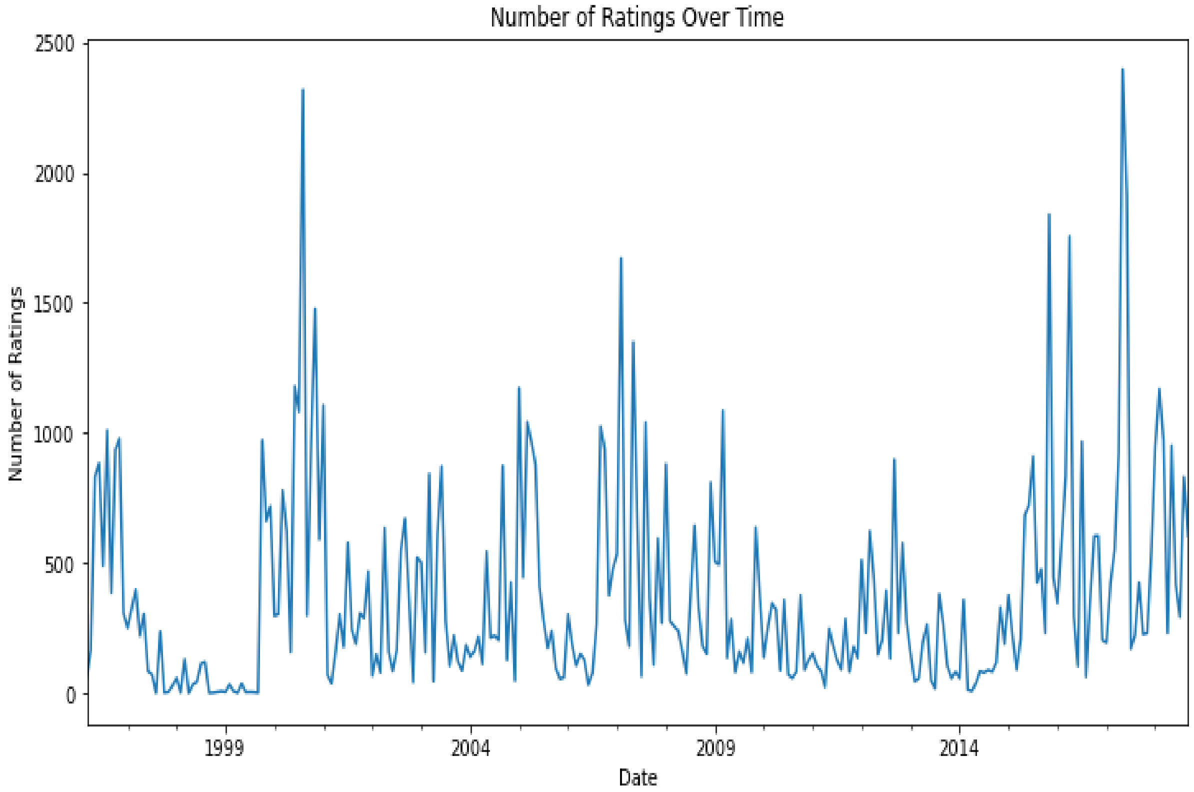

Crime, Mystery, Drama, War, Film-Noir and Documentary are the highest rated genres hence movie producers can prioritize them when producing a movie.

# FINDINGS




Forrest Gump(1994) is the most rated movie with over three hundred ratings.

# FINDINGS




Based on this time series visualization there is no clear pattern for ratings over time.

# METRICS FOR DIFFERENT MODELS

| METRICS | KNN BASIC | KNN BASELINE | KNN MEANS | SVD |
|---|---|---|---|---|
| RMSE | 0.9723 | 0.87859 | 0.9001 | 0.8744 |
| MAE | 0.7488 | 0.6725 | 0.6887 | 0.6719 |

# METRICS FOR DIFFERENT MODELS

- SVD performs the best with the lowest RMSE of 0.8744, indicating it has the most accurate predictions in terms of minimizing error.
- KNNBaseline is a close second with an RMSE of 0.87859, which is still very good and competitive with SVD.
- KNNWithMeans performs moderately well with an RMSE of 0.9001, better than KNNBasic but worse than the other models.
- KNNBasic has the highest RMSE of 0.9723, suggesting it is the least accurate model among the ones compared.

# RECOMMENDATIONS

- ❖ Use hybrid recommendation approach that combines collaborative filtering (e.g., KNN, SVD) with content-based filtering.

- ❖Incorporate user-generated tags to enrich the content-based filtering component. Analyze tag frequencies and their relationships with genres to create more nuanced features.

- ❖Implement a system to gather user feedback on recommendations (e.g., thumbs up/down, ratings). Use this data to refine models and adapt to changing user preferences. Build dynamic user profiles that evolve based on user interactions over time.