Movie Recommendation for "Moviefix" Stakeholder

BY STEPHEN NDIRANGU

Business Understanding

▶ The movie industry is fast growing, and with so many options, there is need for cutting edge user interactivity. This involves tailoring their movie watching experience via intuitive recommendations that can give them interesting options based on what they prefer, while also enticing them with other content that they may like

▶ It is for this very reason we are working on a recommendation system for the "Moviefix" firm, in order to not only capture the users, but also keep them coming back for more. To accomplish this, our model will create 5(five) recommendations that will be given to the user.

Problem Statement

- The main challenge is to design and implement a movie recommendation system that employs collaborative filtering techniques to predict movie preferences for users based on their past ratings. In short, to analyse their past activities and give them recommendations based on their tastes.

- To address a potential "cold start" problem (where new users or movies have limited ratings), if possible, we will attempt to explore a hybrid approach that combines collaborative filtering with content-based filtering.

- The success of this project will be measured by evaluating the accuracy and relevance of the recommendations via metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) that will gauge the performance of the model.

- Ultimately, the objective is to build a recommendation system that enhances user engagement, encourages exploration of diverse movies, and contributes to the overall satisfaction of users on the platform.

Data understanding

The dataset describes 5-star rating and free-text tagging activity from **MovieLens**, a movie recommendation service.
It contains 100836 ratings and 3683 tag applications across 9742 movies. These data were created by 610 users between March 29, 1996 and September 24, 2018.

This dataset was generated on September 26, 2018.

Users were selected at random for inclusion. All selected users had rated at least 20 movies.
No demographic information is included. Each user is represented by an id, and no other information is provided.

The data are contained in the files links.csv, movies.csv, ratings.csv and tags.csv.
The dataset was provided by MovieLens, a movie recommendation service.
It includes the movies and the rating scores made for these movies contains. It contains 100,000 ratings (1–5) from 943 users on 1682 movies.
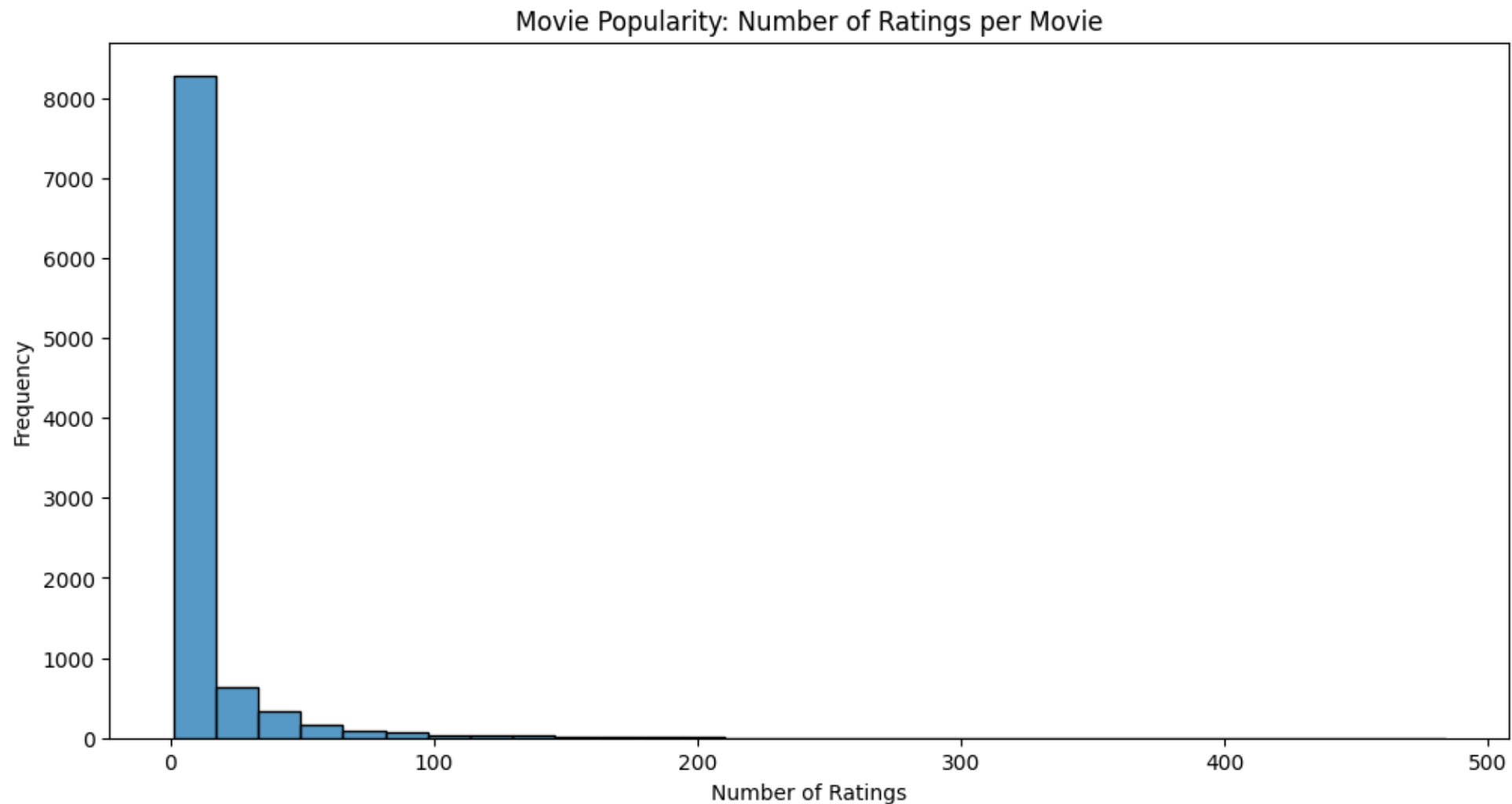
EDA: Genres

Distribution of Movie Genres
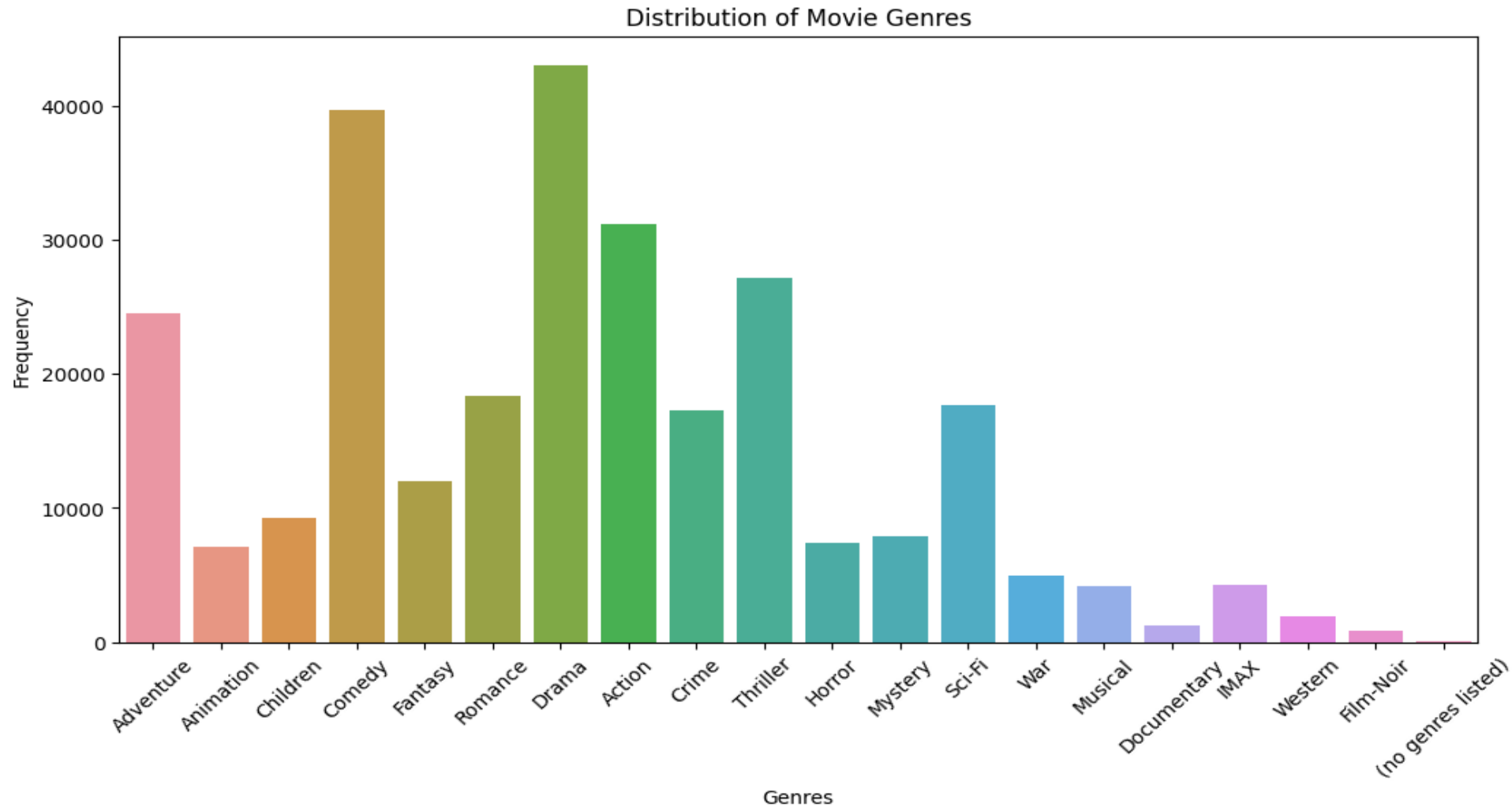
EDA: Ratings Distribution

Distribution of Ratings

EDA: Ratings Per Movie
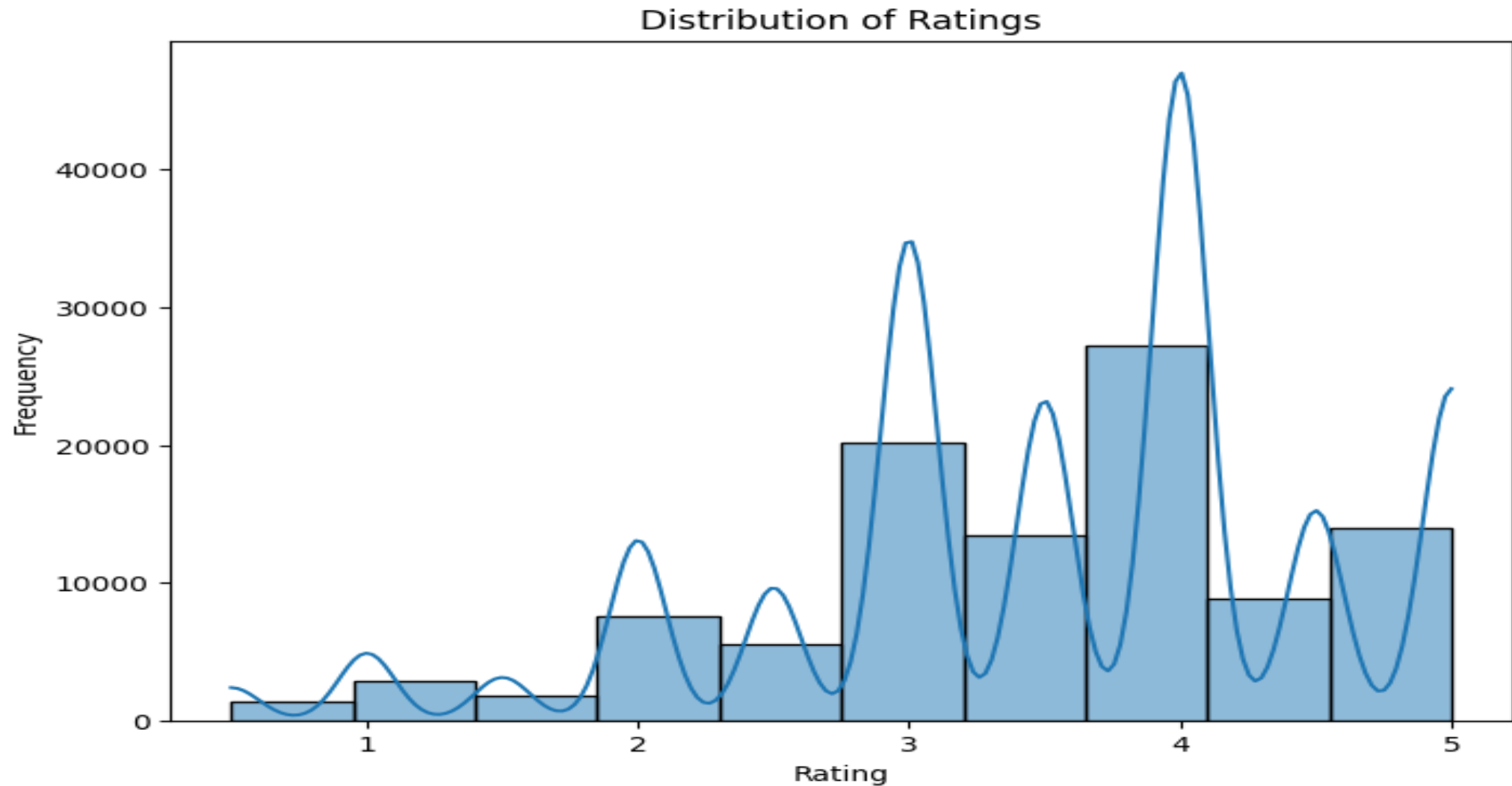
Movie Popularity: Number of Ratings per Movie

Collaborative Filtering

title	userId 1	2	3	4	5	6	7	8	9	10	...	601	602	603	604	605	606	607	608	609
10 Things I Hate About You (1999)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	...	0.000	0.000	3.000	0.000	5.000	0.000	0.000	0.000	0.000
12 Angry Men (1957)	0.000	0.000	0.000	5.000	0.000	0.000	0.000	0.000	0.000	0.000	...	5.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2001: A Space Odyssey (1968)	0.000	0.000	0.000	0.000	0.000	0.000	4.000	0.000	0.000	0.000	...	0.000	0.000	5.000	0.000	0.000	5.000	0.000	3.000	0.000
28 Days Later (2002)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	...	0.000	0.000	0.000	0.000	0.000	0.000	0.000	3.500	0.000
300 (2007)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	3.000	...	0.000	0.000	0.000	0.000	3.000	0.000	0.000	5.000	0.000
40-Year-Old Virgin, The (2005)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	...	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A.I. Artificial Intelligence (2001)	0.000	0.000	0.000	0.000	0.000	0.000	4.500	0.000	0.000	0.000	...	0.000	0.000	0.000	0.000	1.000	3.500	0.000	4.500	0.000
Abyss, The (1989)	4.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	...	0.000	0.000	1.000	0.000	0.000	0.000	0.000	3.000	0.000

Quiz show recommendations

```
recommendations for Quiz Show (1994):

1:Crimson Tide (1995), with distance of 0.47089748328125036:
2:Dances with Wolves (1990), with distance of 0.49850202326272:
3:Get Shorty (1995), with distance of 0.505437474589245:
4:Fugitive, The (1993), with distance of 0.510206572438254:
5:Firm, The (1993), with distance of 0.51172779329904:
```

Toy story Recommendations

```
Recommendations for Toy Story (1995):

1: Toy Story 2 (1999), with distance of 0.427398739680284
2: Jurassic Park (1993), with distance of 0.434363195913843
3: Independence Day (a.k.a. ID4) (1996), with distance of 0.435738306472334
4: Star Wars: Episode IV - A New Hope (1977), with distance of 0.442611829420063
5: Forrest Gump (1994), with distance of 0.452904092059826
```

Recommendations

▶ **Hybrid Recommendations**: While not implemented in this project, combining collaborative filtering and content-based recommendations can enhance the accuracy and coverage of recommendations. Hybrid models address the limitations of each approach, providing more robust suggestions for users.

▶ **Fine-Tuning and Evaluation**: Experiment with different hyperparameters, algorithms, and preprocessing techniques to fine-tune the recommendation models further. Utilize evaluation metrics like RMSE, precision, recall, and F1-score to objectively assess the models' performance.

▶ **User Interface and Deployment** : Consider creating a user interface where users can input their preferences and receive personalized recommendations. This interface could be deployed on a web server, allowing users to interact with the recommendation system seamlessly.

Conclusion

▶ In this project, we set out to build a movie recommendation system using collaborative filtering and content-based approaches. We explored a diverse dataset containing movie ratings, genres, and user interactions to create a personalized movie recommendation system. Through data cleaning, exploratory data analysis (EDA), and the implementation of recommendation algorithms, we've gained insights into the movie preferences of users and successfully generated movie recommendations.

Thanks