PROJECT PRESENTATION

MOVIE RECOMMENDATION SYSTEM

DATA SCIENTIST INCLUDE:

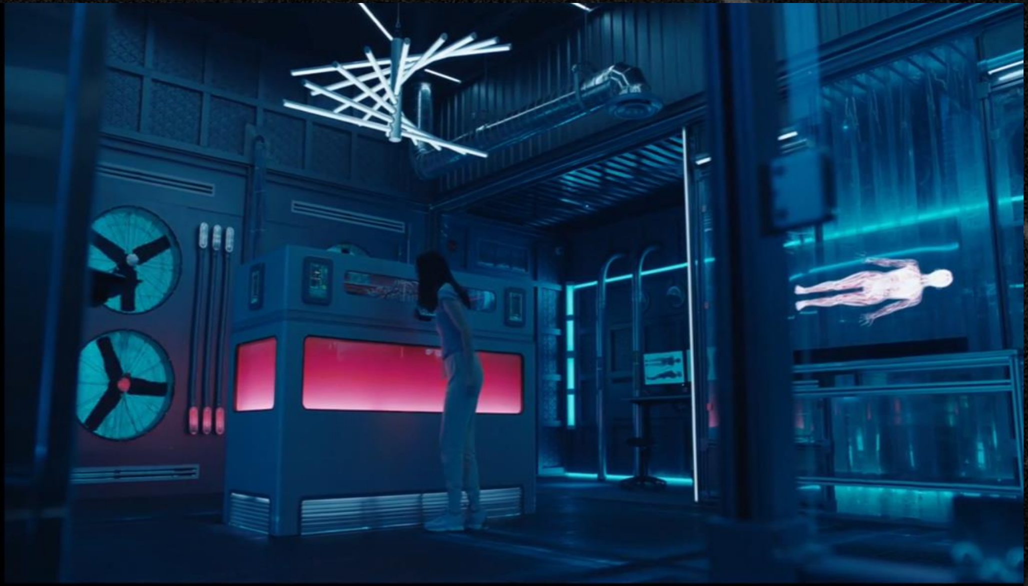
Isoe Calmar
Karue Daniel
Pamelah Veneser
Kalia Precious
Liz Sharleen

THE AIM OF THE PROJECT

This project aims to develop a recommendation system that provides personalized movie suggestions to users based on their past ratings, and recommend the top 5 movies that a user is likely to enjoy.

STAKEHOLDERS

The stakeholders for this project are a streaming platform aiming to enhance their services and retain viewers by offering personalized recommendations and improving content acquisition strategies. This project will help them achieve that, by analyzing user preferences, identifying popular genres, and optimizing recommendations to increase user engagement and satisfaction.



BUSINESS PROBLEM

This project aims to develop a machine learning-based recommendation system that provides personalized movie suggestions by analyzing users' past ratings, thereby improving their overall experience and engagement.

OUR OBJECTIVES

- Develop a recommendation system that provides personalized movie recommendations based on user ratings.

- Implement collaborative filtering and content-based approaches to improve recommendation accuracy.

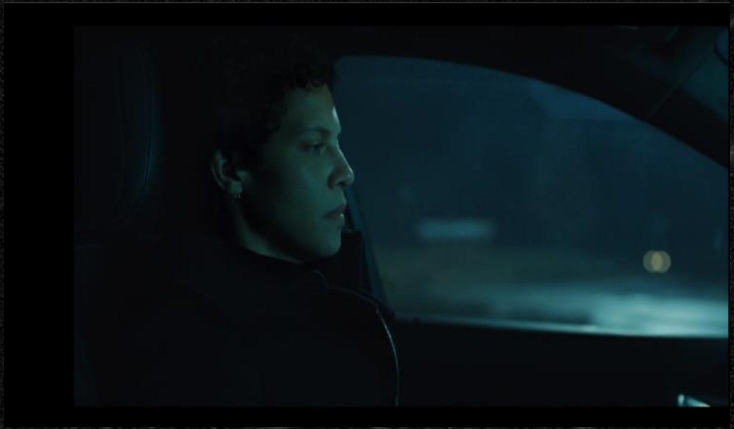
- Evaluate the recommendation system using appropriate performance metrics to ensure relevance and accuracy.

- Provide actionable insights to stakeholders to enhance user satisfaction and engagement strategies.

DATA SOURCE

The dataset used in this project is sourced from the GroupLens Research team, titled "MovieLens Latest Small Dataset." It is a CSV-based dataset containing user ratings, movie information, and additional features relevant for building a recommendation system. You can access it here: [MovieLens Dataset](https://grouplens.org/datasets/movielens/latest/).

The data is contained in the files `links.csv`, `movies.csv`, `ratings.csv` and `tags.csv`.



GENERAL WORKFLOW

BUSINESS UNDERSTANDING

↓

DATA UNDERSTANDING

↓

DATA PREPARATION

↓

MODELING

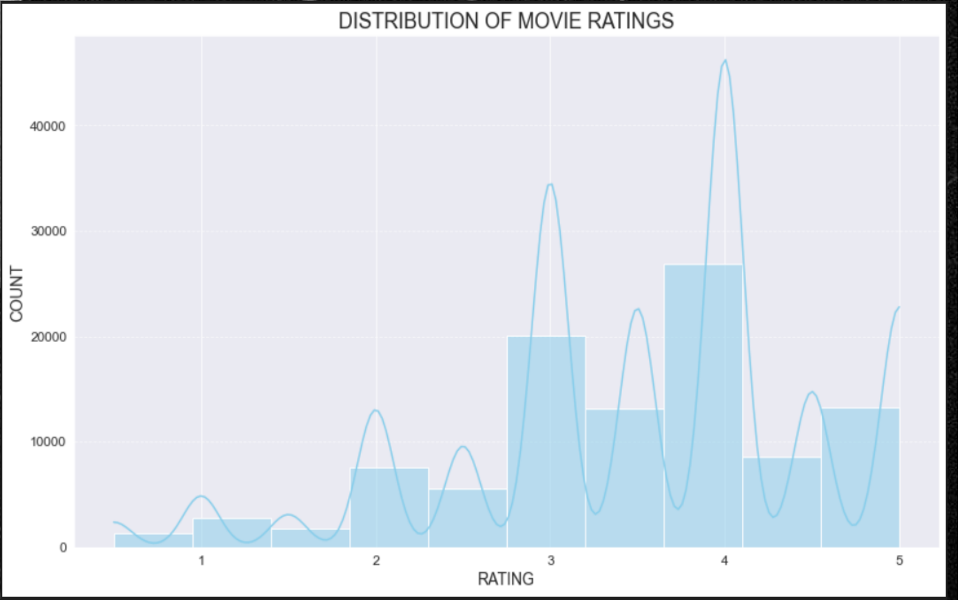
↓

EVALUATION

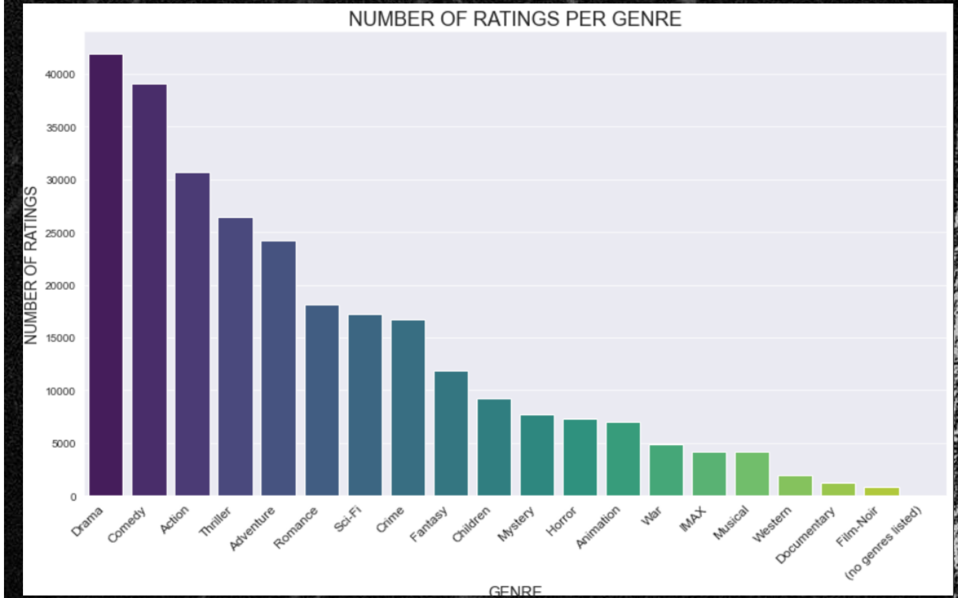
↓

CONCLUSION AND RECOMMENDATION

1. DATA VISUALIZATION



Here is a histogram showing how ratings are distributed across the dataset to identify trends like whether users tend to give higher or lower ratings.

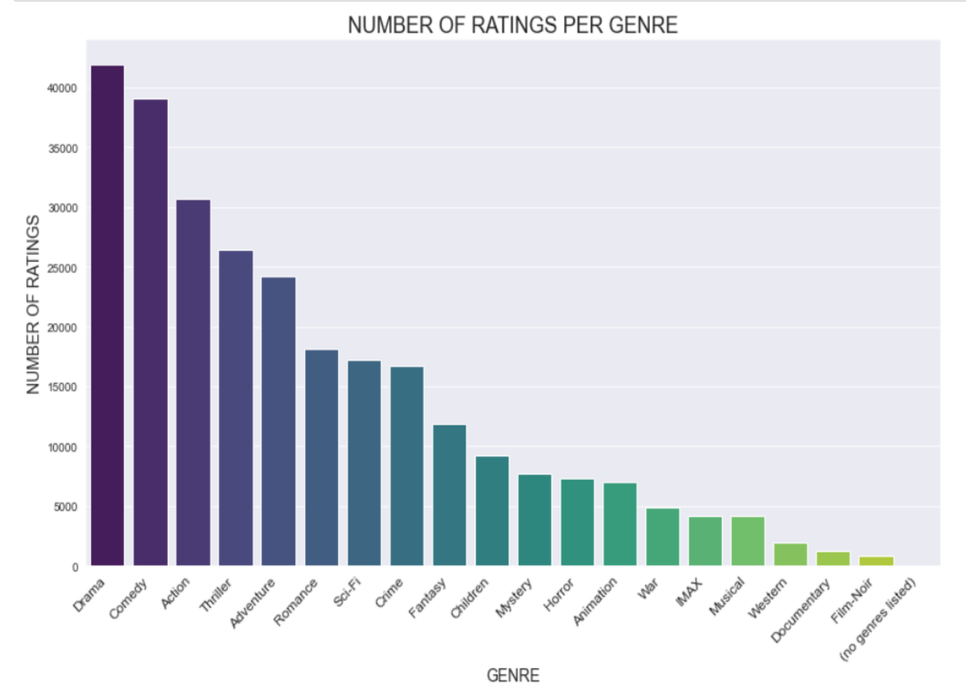


Here is a Barplot showing the number of ratings per genre

GENRE POPULARITY ANALYSIS

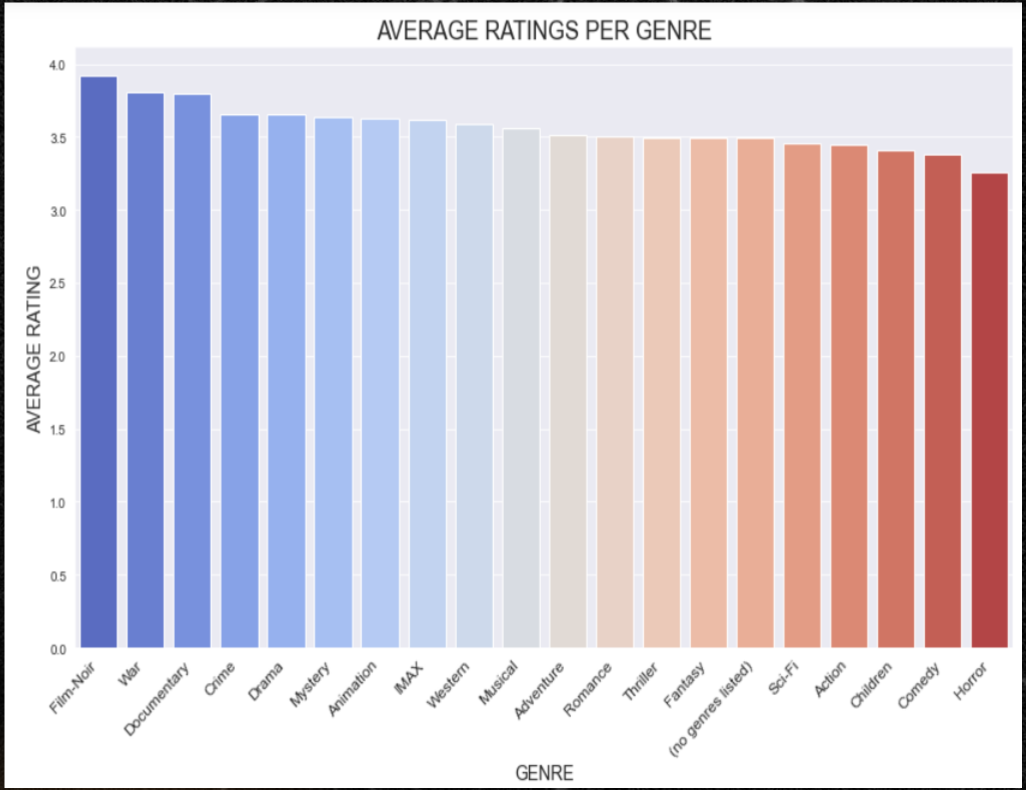
This plot shows the number of ratings across different movie genres, with some genres receiving significantly higher engagement than others.

As observed, genres like drama, comedy action, thriller, and adventure are the most-rated genres while genres like film-noir, musicals, documentary, and western are some of the least-rated genres.

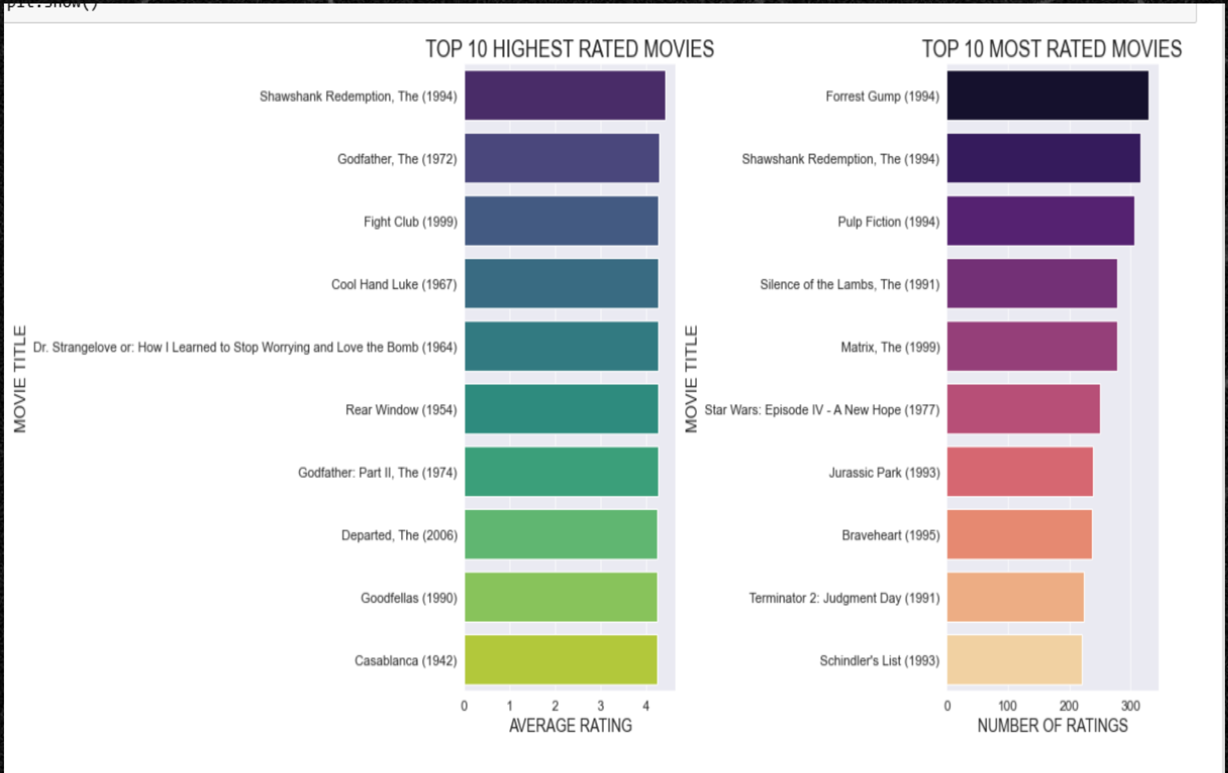


This plot highlights the average rating for each genre, showing **variations in user satisfaction**.

As observed, despite being among the least-rated genres, film-noir received the highest average rating, followed closely by war and documentary genres



MOST-RATED AND TOP-RATED MOVIES



Observation

The visualization reveals that certain movies have received significantly higher average ratings, with the highest-rated movies having an average rating above 4.5 . However, as observed in the plots above, these movies might have a lower number of ratings overall. As mentioned earlier, this might indicate a niche but satisfied audience.

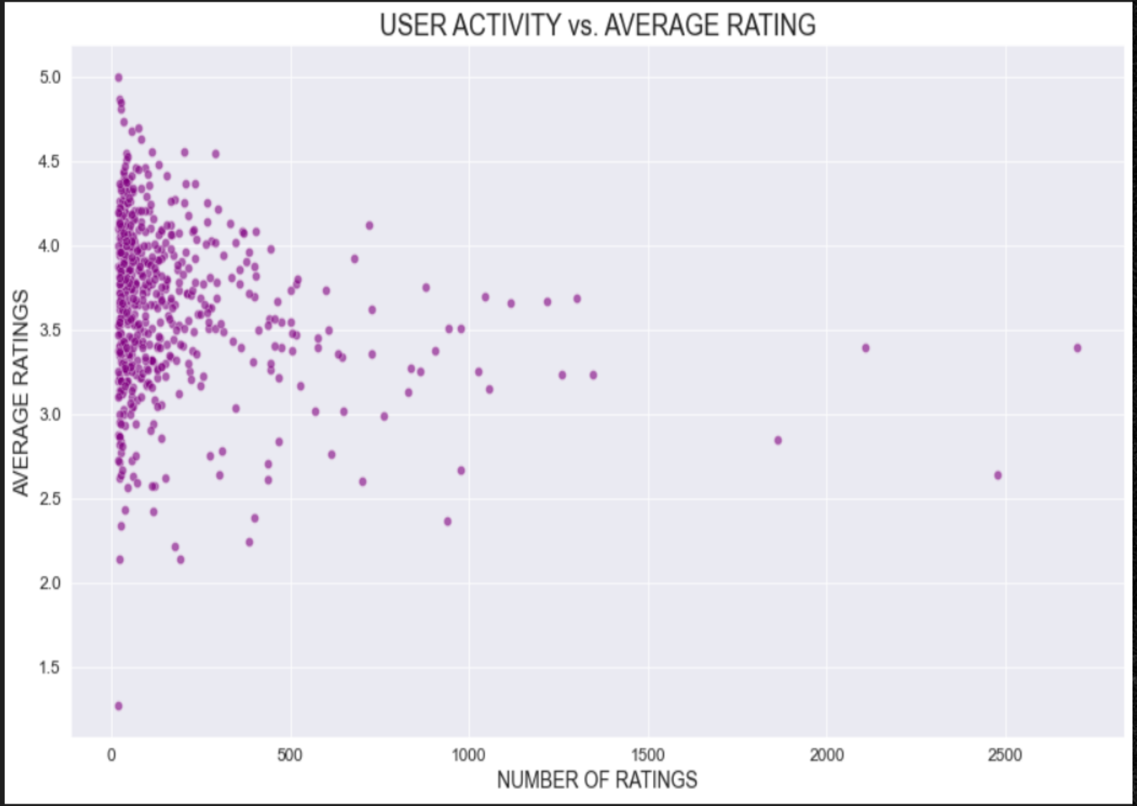
On the other hand, the most-rated movies suggest broader audience engagement, indicating their widespread popularity but with slightly lower average ratings.

Movies like The Shawshank Redemption(1994) appear both in the top-rated and most-rated list suggesting a very popular movie that was also did well.

CORRELATION ANALYSIS BETWEEN USER ACTIVITY AND RATINGS

This scatter plot suggests that the most active users tend to have average ratings within a specific range, whereas less active users show more variability in their ratings.

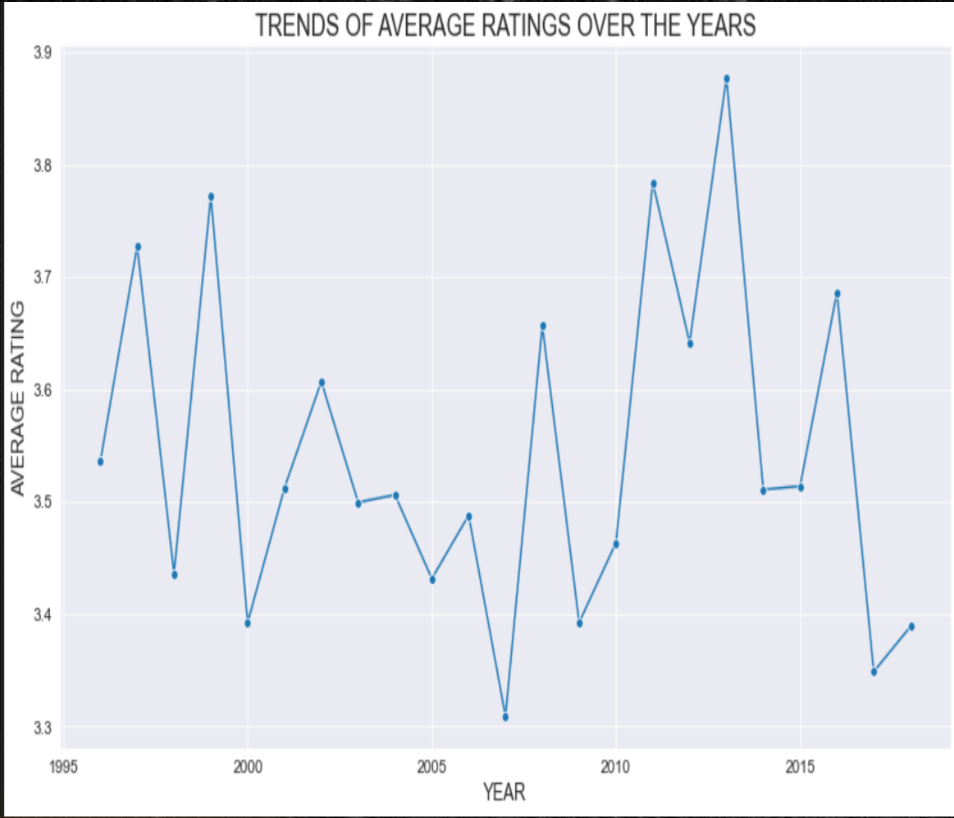
Highly active users might have more consistent preferences compared to sporadic users.



RATING TRENDS OVERTIME

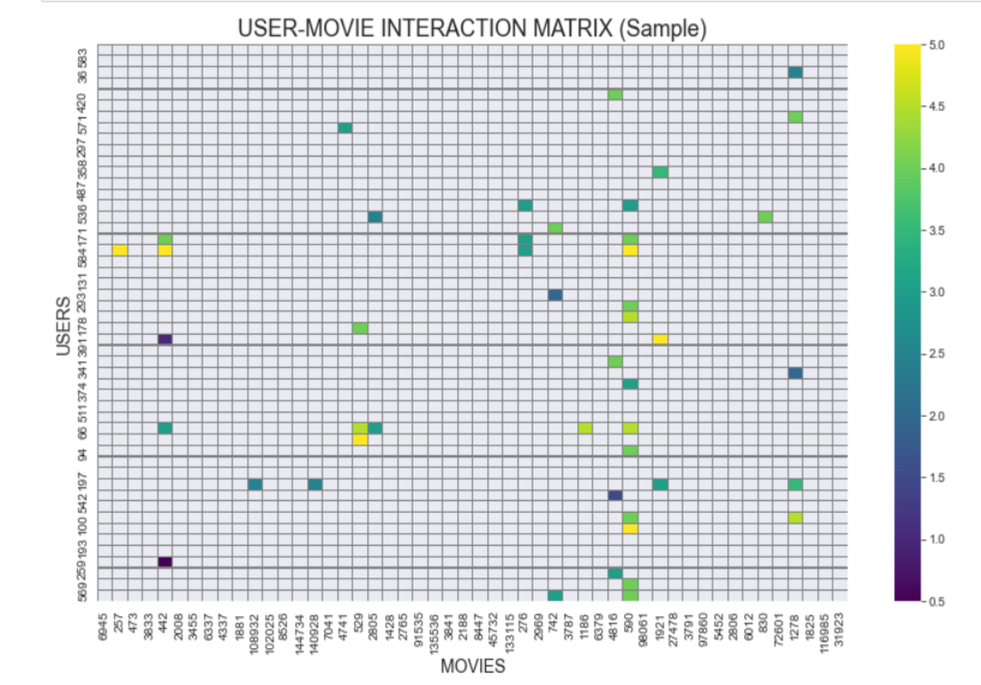
From this lineplot we observe that the average rating of movies has fluctuated over the years with a noticeable decline around the years 2006-2007.

This fluctuations could potentially be influenced by factors such as changes in user preferences, movie quality over time, or other factors such as the rise of social media.



Visualizing sparcity in ratings

This visualization demonstrates that the user-movie interaction matrix is **highly sparse**, with most cells having missing values (no rating). This sparsity indicates that user's rate only a small fraction of available movies, which is typical in such a dataset.



CONCLUSION

Based on the analysis and modeling conducted on the MovieLens dataset, we've gained valuable insights into user preferences, rating behaviors, and the effectiveness of recommendation models. Our exploratory data analysis highlighted key patterns in user engagement, genre preferences, and rating distributions. The recommendation system, optimized using collaborative filtering techniques, achieved a promising RMSE score indicating a reasonable level of accuracy. These findings provide actionable strategies to enhance movie recommendations, user experience, and business growth. Future improvements can further optimize the system to address challenges such as cold start problems and evolving user preferences.

RECOMMENDATIONS

1. **Personalized promotion strategies:**

   Focus marketing efforts on movies consistently rated above 4 and promote highly rated niche genres to attract dedicated audiences.

2. **Genre-based content expansion:**

   Increase content acquisition and production in high-demand genres like drama, comedy, and action to align with user preferences.

3. **User segmentation for targeted engagement:**

   Develop loyalty programs and premium recommendation services for highly active users who provide more stable and valuable feedback.

4. **Peak engagement planning:**

   Leverage historical rating trends to schedule content releases and promotions during peak engagement periods to maximize impact.

5. **Differentiated recommendation strategies:**

   Offer separate recommendation lists for "popular" and "high-quality" movies to cater to diverse user preferences and expectations.

6. **Cold-start mitigation:**

   Introduce content-based filtering or hybrid models to enhance recommendations for new users with limited rating history.

NEXT STEPS

- Build a hybrid recommendation model: Combine collaborative and content-based filtering to improve accuracy and address limitations of individual approaches.

- Expand feature engineering: Incorporate additional user and movie features, such as demographics, review text, and browsing history, to enrich the recommendation model.

- Deploy and monitor the system: Implement the model in a real-world environment, track its performance over time, and continuously refine it based on user feedback and new data.

- Collect more data that is also diverse: Collect additional data to improve the model robustness.