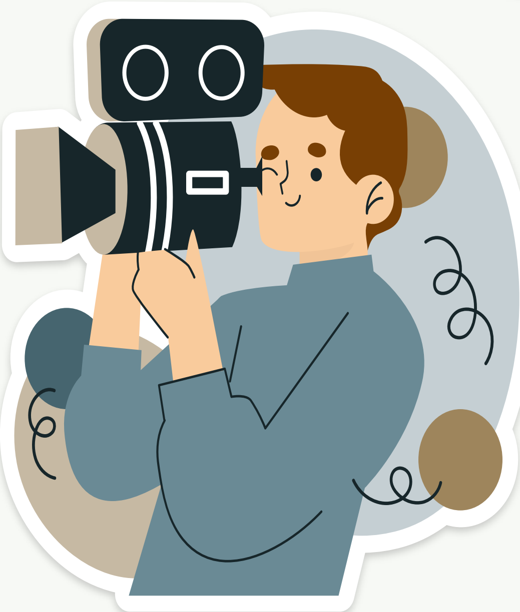
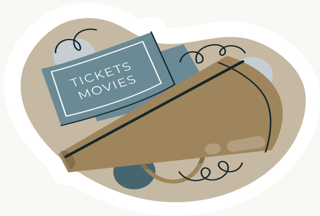
MOVIE RECOMMENDATION

SYSTEM

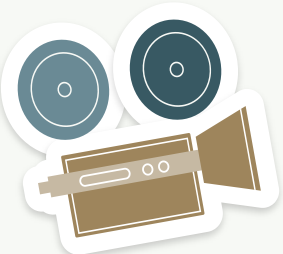
GROUP MEMBERS:

1. Randell Mwania
2. Priscilla Wairimu
3. Joy Wangui
4. Simon Ng’ethe
5. Julius Kagety
6. Geoffrey Rotich



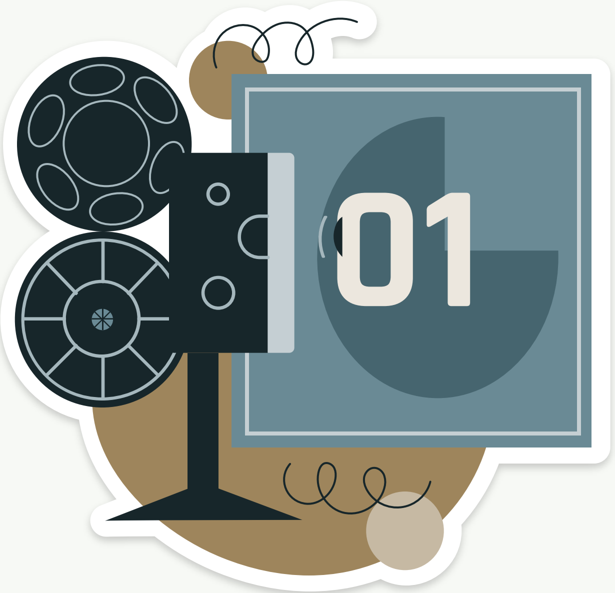


Contents



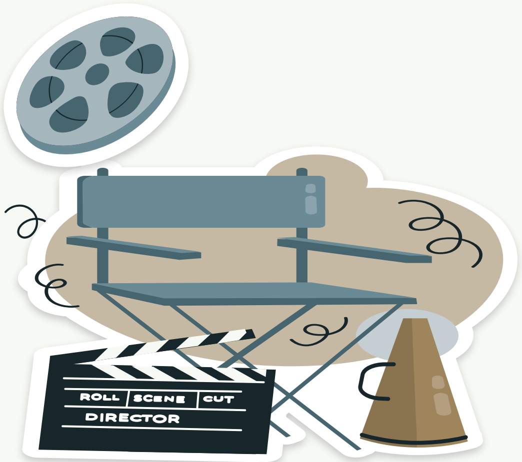
BUSINESS OVERVIEW

This project crafts a movie recommendation system to heighten user satisfaction in the entertainment industry. By leveraging historical data, it tailors personalized movie suggestions to foster engagement and drive user retention.



02 BUSINESS UNDERSTANDING

2.1 OBJECTIVES



Main objective:

- To provide personalized movie recommendations tailored to individual user preferences.

Specific objectives:

- To find the most popular movies
- To build a content based recommendation system
- To build a collaborative filtering movie recommendations based on user interactions.

2.2 PROBLEM STATEMENT

Choosing movies can be overwhelming, leading to decision fatigue and less engagement. This project tackles the issue with a personalized recommendation system, using user data to make movie exploration easier and more enjoyable, improving satisfaction and retention.

(03) DATA UNDERSTANDING

For this project, the dataset is sourced from Kaggle, available here. It comprises three key files:

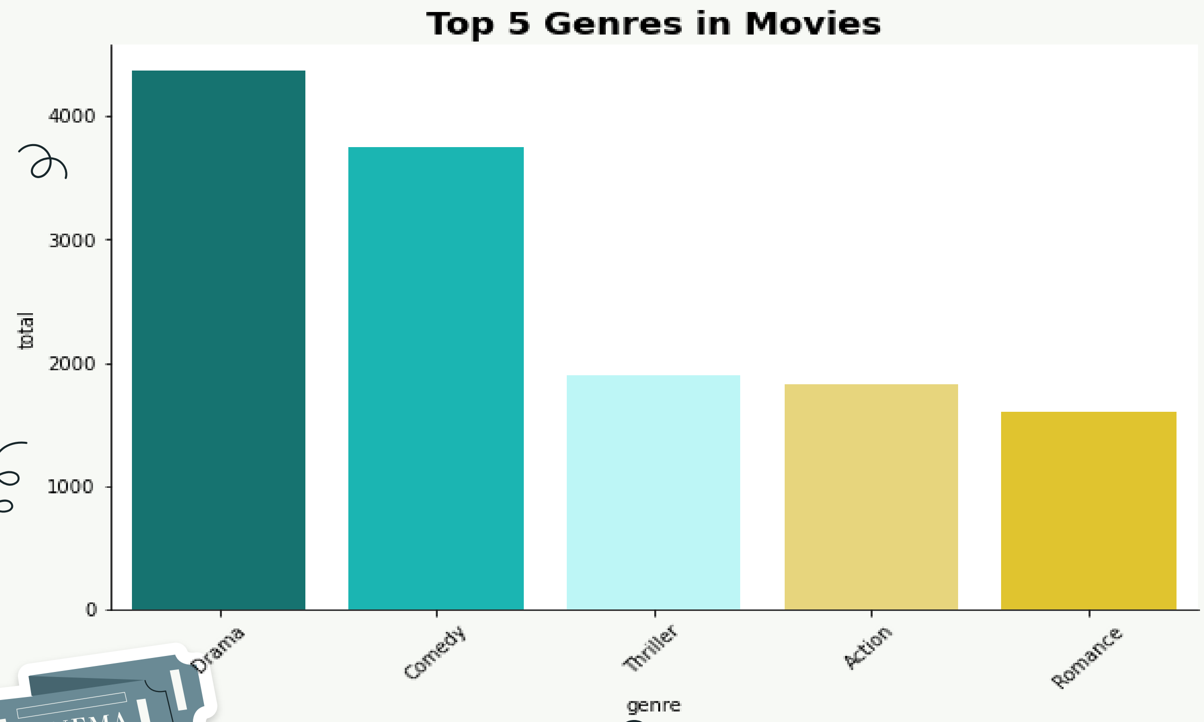
tag.csv: Users apply tags to movies, with data including user ID, movie ID, tag, and timestamp.

rating.csv: This file contains user ratings for movies, including user ID, movie ID, rating, and timestamp.

movie.csv: Information on movies is stored here, detailing movie ID, title, and genres.

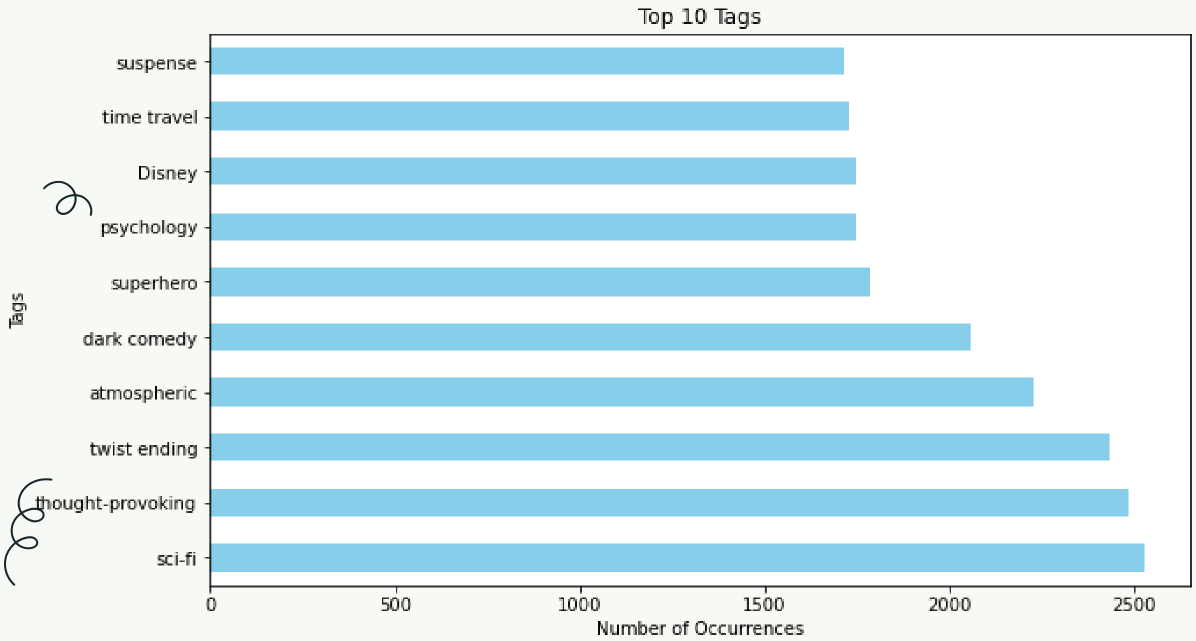
The dataset has 233213 rows and 9 columns.

(04)DATA ANALYSIS



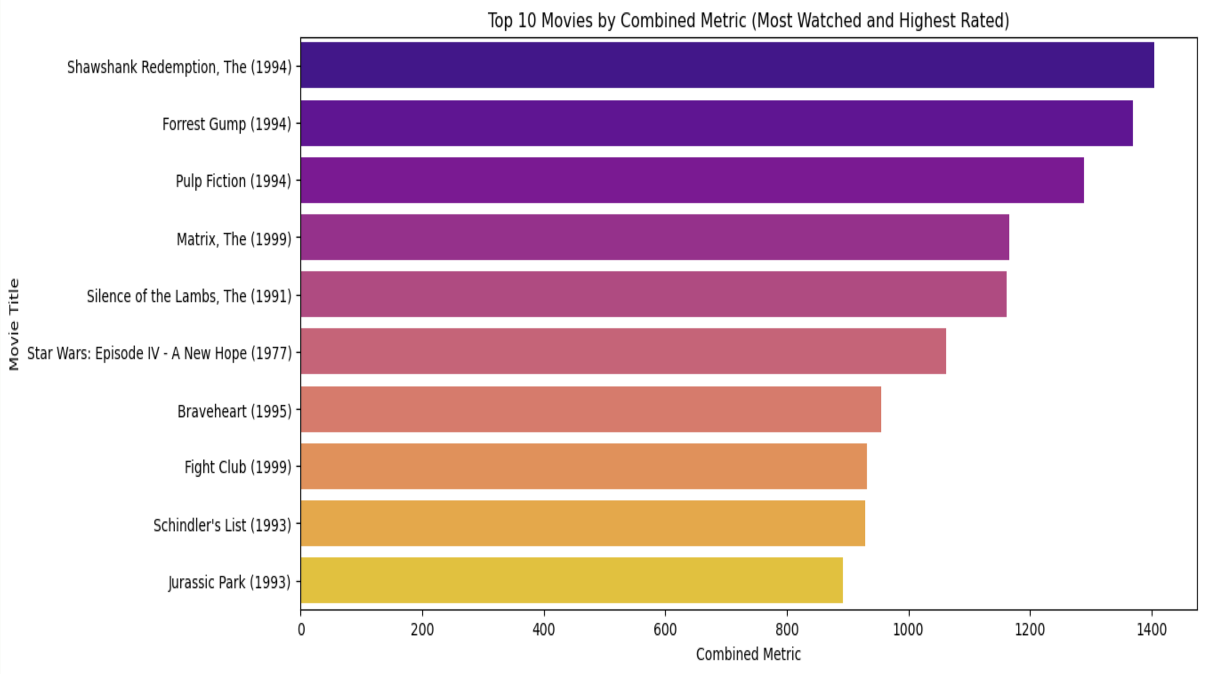
From the visualization of the Top 5 movie genres, it is evident that Drama was the top most movie genre among users.

(04)DATA ANALYSIS

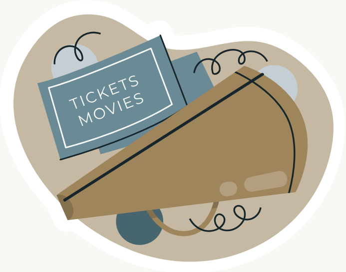


This shows the top 10 movie tags. Sci-fi was the most common tag followed closely by thought provoking and twist ending.

(04)DATA ANALYSIS



This shows the top 10 movie tags watched in relation to the movie rating. ShawShank Redemption was the top movie with a rating of 1404 votes.



MODELING

Collaborative Filtering using Surprise library

This approach utilizes the Surprise library for collaborative filtering.
The results were as follows:

Evaluation metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std deviation
RMSE(Test set)	0.8733	0.8795	0.8728	0.8707	0.8739	0.874	0.0029
MAE (Test set)	0.673	0.6766	0.699	0.6696	0.66701	0.6717	0.0028
Fit time	6.47	6.63	6.67	6.45	6.69	6.58	0.1
Test time	0.34	0.2	0.31	0.2	0.31	0.27	0.06

INTEPRETATION OF RESULTS

RMSE (Root Mean Squared Error):
RMSE is a measure of how well the model predicts the ratings.
Lower RMSE values indicate better predictive performance.
In the results; The mean RMSE across folds is 0.8740.

MAE (Mean Absolute Error):
MAE measures the average absolute difference between predicted and actual ratings.
Like RMSE, lower MAE values indicate better performance.
In the results; The mean MAE across folds is 0.6717.

Fit Time and Test Time:
Fit Time represents the time taken to train the model while test time represents the time taken to test the model on the test set.
Lower fit times are generally preferred for efficiency while lower test time is preferred for predictions.

MODELING

Content-Based Filtering:

This approach uses TF-IDF vectorization and cosine similarity for content-based filtering. Content-based filtering is a recommendation system technique that suggests items as movies to users based on the characteristics of items and the preferences expressed by the user. The primary idea is to recommend items that are similar to those the user has liked or interacted with in the past.

Hybrid Recommendation System (Collaborative + Content-Based)

This approach integrates collaborative and content-based filtering for a hybrid recommendation system. The hybrid system seamlessly integrates both collaborative and content-based filtering. When a user seeks recommendations, the system dynamically decides whether to emphasize collaborative insights or content characteristics.

(06) CONCLUSION

- **Movie Recommendation System**:

The project focuses on building a movie recommendation system to address the challenge of information overload in the face of numerous movie choices.

- **Content-Based and Collaborative Filtering:**

Content-Based Filtering uses TF-IDF vectorization and cosine similarity to suggest movies based on user preferences and item characteristics.

Collaborative filtering relies on user-item interactions, recommending items based on patterns and preferences observed by similar users.

- **Hybrid Recommendation System:**

The hybrid recommendation system integrates both collaborative and content-based filtering. It dynamically selects the approach based on user input, offering a well-rounded and personalized recommendation experience.

- **Evaluation Metrics:**

The collaborative filtering model using the Surprise library is evaluated using RMSE and MAE metrics across multiple folds, showcasing its predictive accuracy and efficiency.

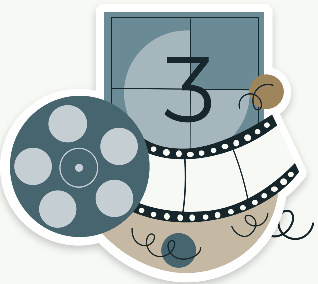
(06) RECOMMENDATIONS

Movie Genres

- The movie streaming sights should focus on production of genres such as drama as it is the most popular.

Movies

- The platform should also consider reboots since most popular movies seem to be older movies over the newer movies.



THANKS!!