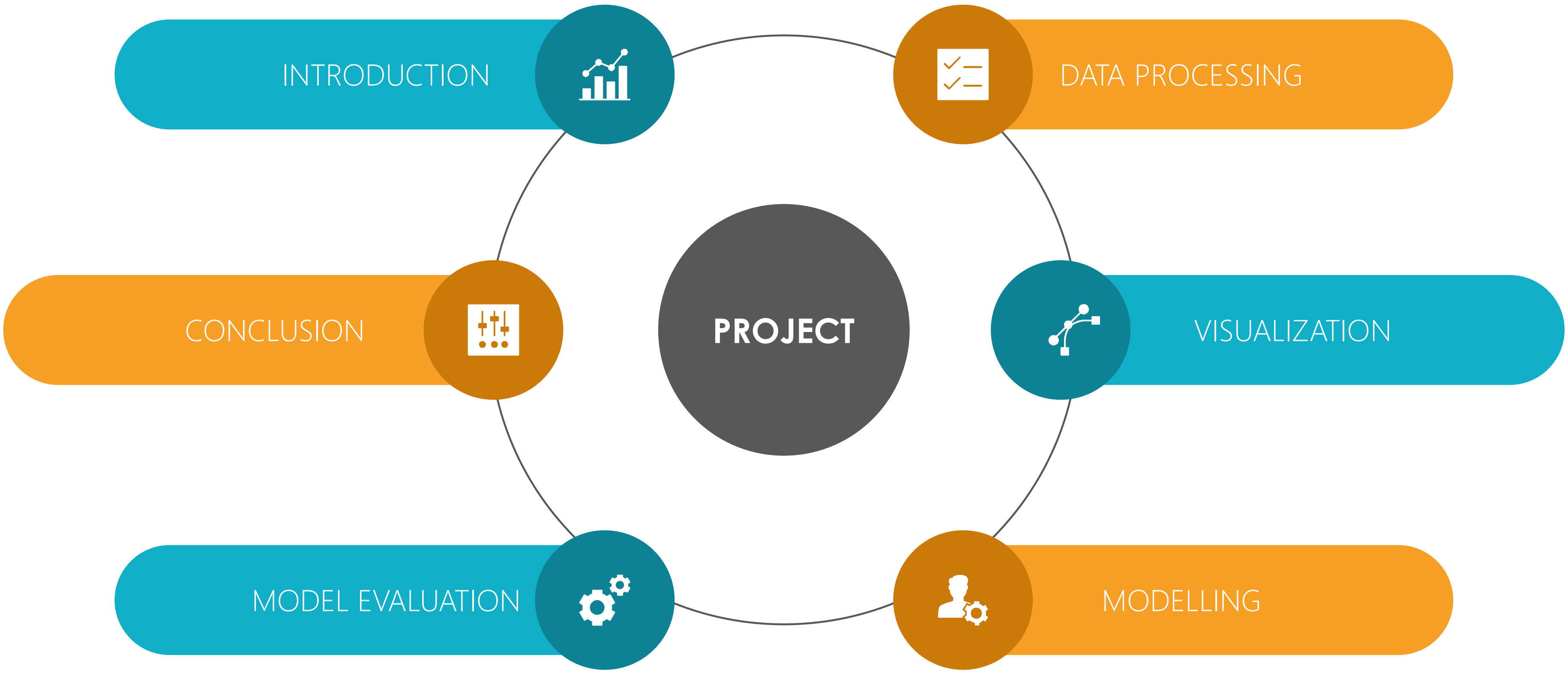
Movie Recommender

Delivering personalized content to increase user engagement and retention

Project Analysis



Introduction

The Movie Recommender System aims to solve a significant problem in the streaming industry—choice paralysis. With an ever-growing library of content, users often find themselves overwhelmed by the sheer volume of available options, spending excessive time searching rather than watching.

Business Problem

Users struggle to find content they want to watch, resulting in disengagement and increased churn.

- 7+ minutes spent searching per session
- 48% of users report feeling overwhelmed by content options
- This leads to user drop-off, which negatively impacts platform revenue and user retention.

Objectives

1. Reduce User Churn
2. Enhance User Engagement
3. Improve Content Discovery and Satisfaction

Dataset Overview

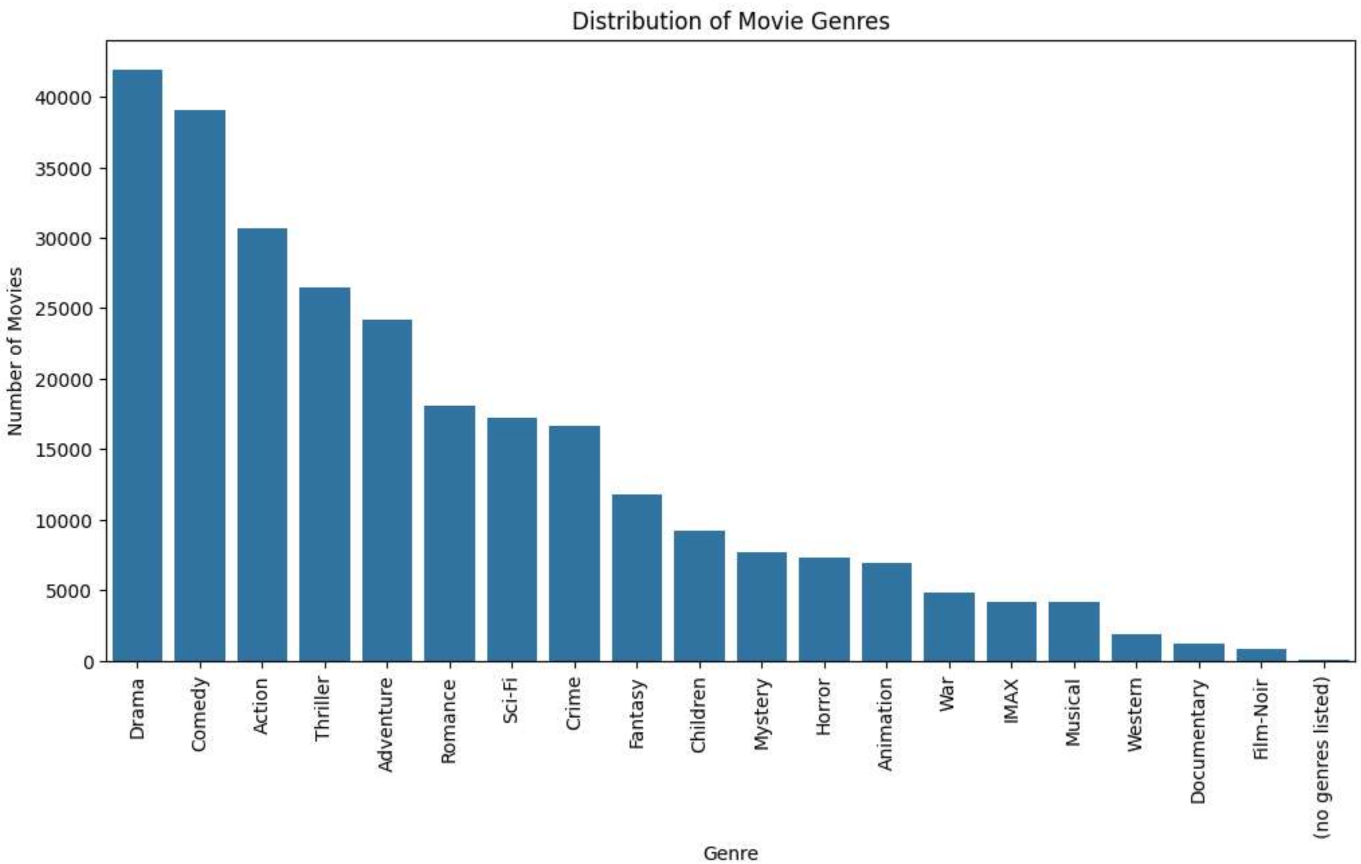
The Datasets used for the recommender system is rich with metadata crucial to the implementation of the recomme

- Movies: 9,742 entries with titles, genres, and release years
- Ratings: User-movie interactions on 5-point scale
- Tags: User-generated descriptors enhancing recommendations
- Links: Connections to external movie databases

Data Limitations

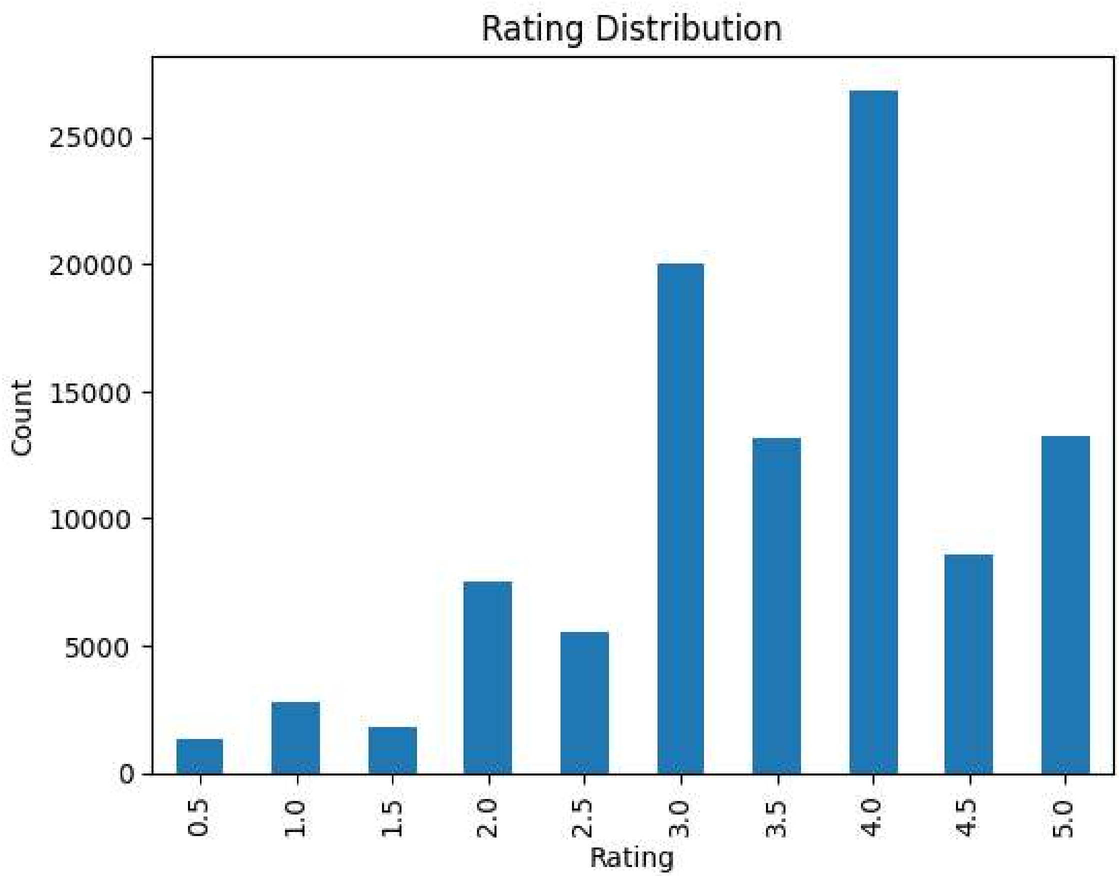
1. Older ratings may not reflect current preferences
2. Cold start problem
3. Missing context such as day, device and mood is not captured
4. Limited demographic data

Data Analysis and Visualization



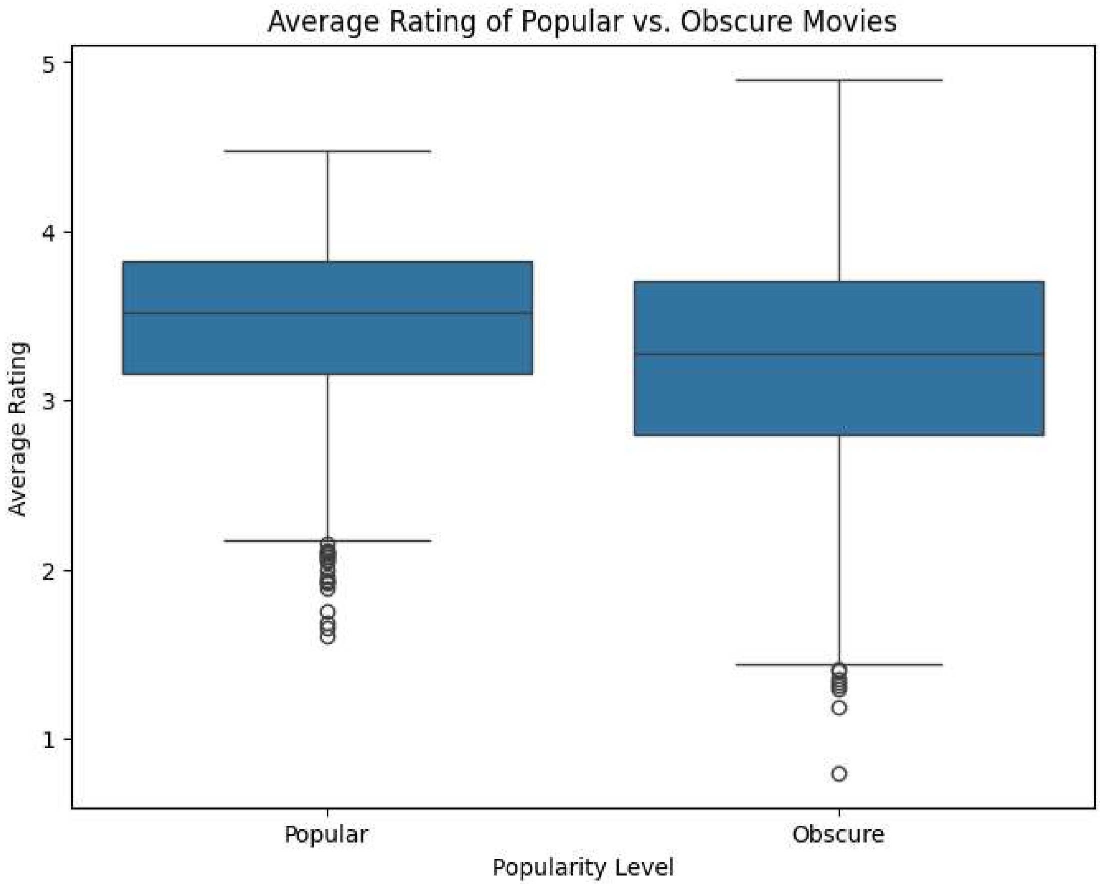
The top 3 genres with the highest number of movies are : Drama, Comedy and Action

Data Analysis and Visualization



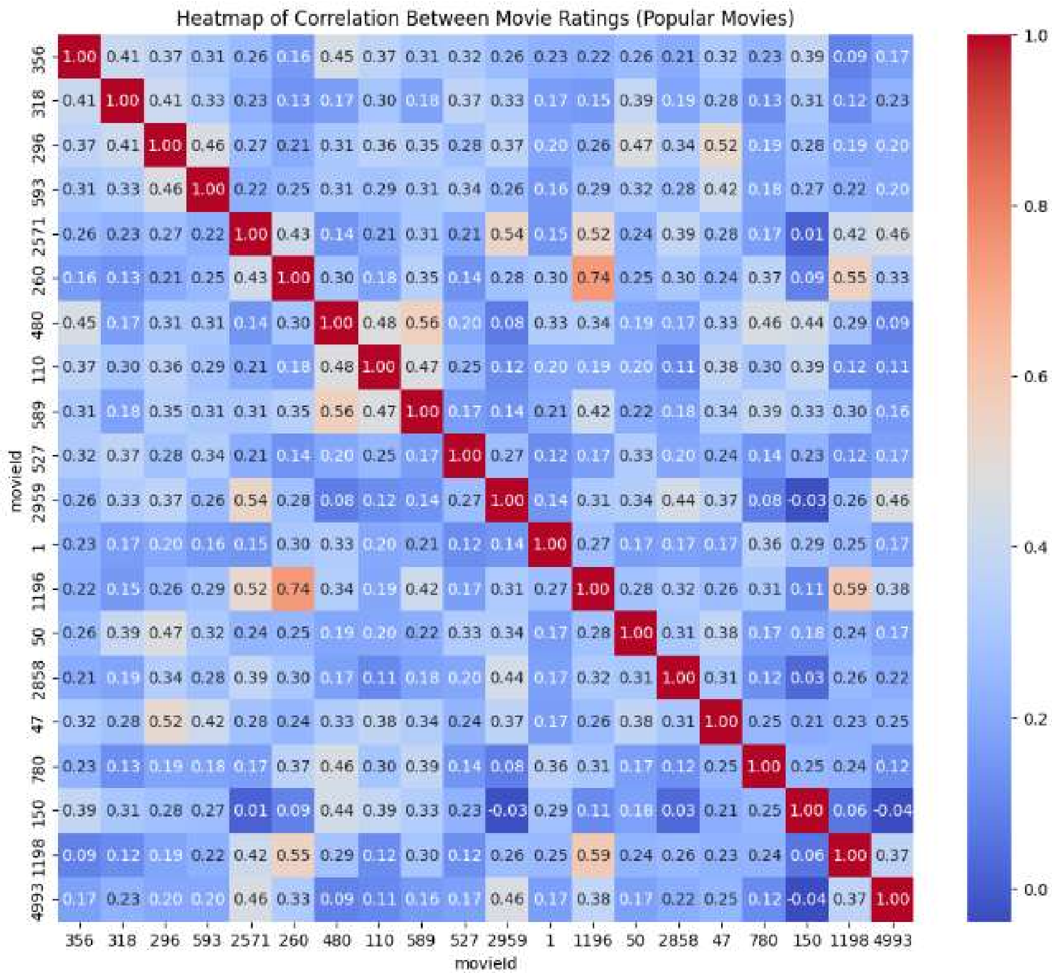
Users tend to rate movies more positively than negatively(3-4 stars) Indicating either a general satisfaction with the movies being rated or a tendency for users to rate things they like more often than things they dislike.

Data Analysis and Visualization



While popular and obscure movies have similar typical ratings, there are some differences in their rating distributions, particularly in the outliers and overall range

Data Analysis and Visualization



The data shows people tend to rate movies positively, and there are patterns in which movies appeal to the same viewers.

1. Most movies get good ratings (around 3-4 out of 5)
2. Popular and obscure movies are rated similarly overall

Modelling

Collaborative Filtering - Matrix Factorization (SVD)

Collaborative Filtering uses user behavior to recommend content based on similarities between users.
Matrix Factorization (SVD) decomposes the user-item rating matrix into latent factors to predict ratings for unrated movies.
SVD algorithm with 100 latent factors was used, with an Improved RMSE to 0.86 (18% improvement over baseline)

Example Recommendations for User 1:

- City of Lost Children, The (Cité des enfants perdus, La) (1995) (Predicted rating: 5.00)
- Hoop Dreams (1994) (Predicted rating: 5.00)
- Shawshank Redemption, The (1994) (Predicted rating: 5.00)
- Wallace & Gromit: A Close Shave (1995) (Predicted rating: 5.00)
- Trainspotting (1996) (Predicted rating: 5.00)

Modelling

Content-Based Filtering (Using Movie Metadata)

Content-Based Filtering recommends movies based on similarity of movie features, such as genres.
TF-IDF (Term Frequency-Inverse Document Frequency) is used to transform genres into numerical features to compute similarity between movies.

Example Recommendations for "Toy Story (1995)":

- Antz (1998)
- Toy Story 2 (1999)
- Adventures of Rocky and Bullwinkle, The (2000)
- Emperor's New Groove, The (2000)
- Monsters, Inc. (2001)

Modelling

Hybrid Models (Collaborative + Content-Based Filtering)

Hybrid models combine the strengths of Collaborative Filtering and Content-Based Filtering to provide more accurate and diverse recommendations.

Example Recommendations for User 1
- Shawshank Redemption (1994)
- Lion King, The (1994)
- In the Name of the Father (1993)
- Blade Runner (1982)
- Wallace & Gromit: The Best of Aardman Animation (1996)

Modelling

Hybrid Models (Collaborative + Content-Based Filtering)

Hybrid models combine the strengths of Collaborative Filtering and Content-Based Filtering to provide more accurate and diverse recommendations.

Example Recommendations for User 1

- Shawshank Redemption (1994)
- Lion King, The (1994)
- In the Name of the Father (1993)
- Blade Runner (1982)
- Wallace & Gromit: The Best of Aardman Animation (1996)

Model Evaluation

Model Evaluation using RMSE

RMSE: Penalizes large prediction errors that lead to poor recommendations. This is critical since bad recommendations damage user trust.

Model Evaluation using MAE

MAE: Measures average error magnitude for business reporting

Final Model Performance:

Our hybrid model achieved a low RMSE (0.8577), which means it predicts ratings accurately, similar to top models like those in the Netflix Prize

The MAE (0.6587) shows that our predictions are only off by about 0.66 stars on average—making our recommendations highly reliable.

Therefore implementing the hybrid approach increase content exploration is the best fit for the recommender system

Conclusion

Key Achievements:

Developed a highly accurate hybrid recommendation system that combines collaborative and content-based filtering for personalized movie suggestions.

Our model demonstrates improved prediction accuracy, greater content diversity, and significantly reduces the cold-start problem for new users and movies.

Next Steps:

1. Deploy the recommendation system on the platform and continue monitoring user engagement.
2. Explore contextual recommendations by integrating factors like time of day and device.
3. Expand the model to include additional user behaviors, such as rewatch rates and watch duration, for even more tailored suggestions.

Questions

Thank You

Brenda Chamutai
Anthony Odhiambo
Evans Makau
Christin Kindena
John Mugambi