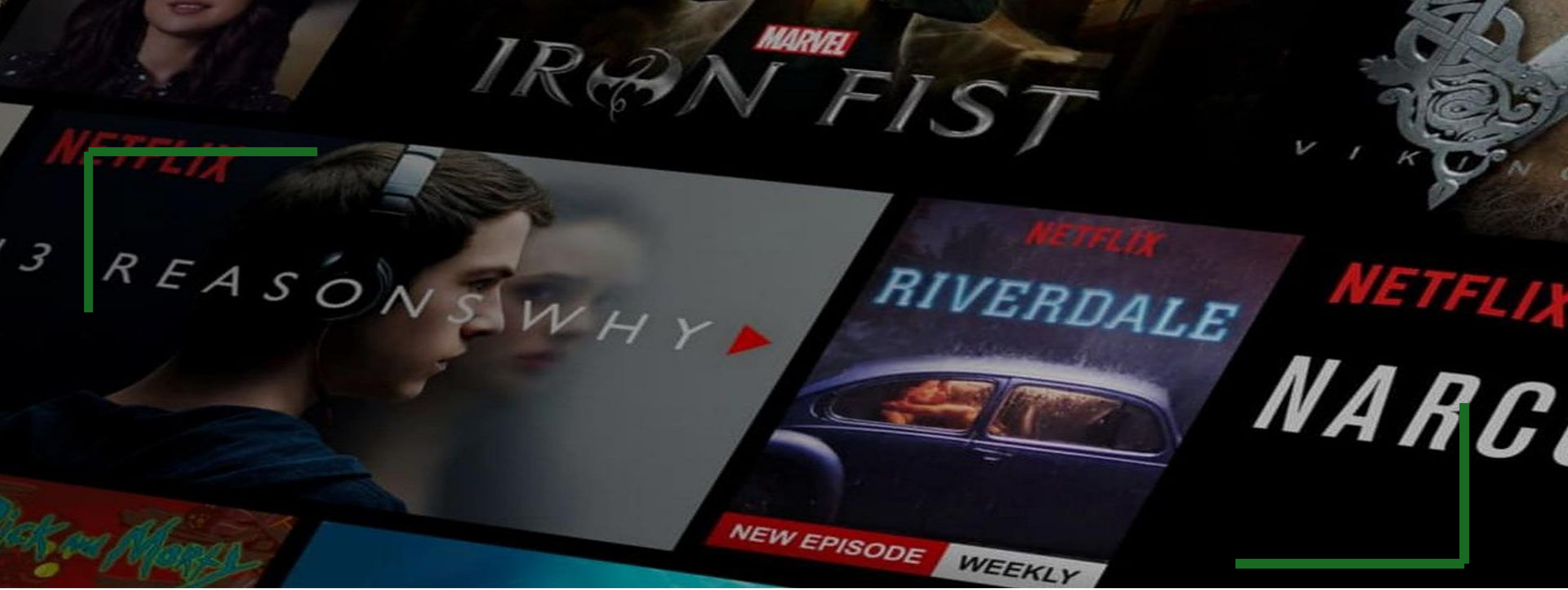


MOVIELENS RECOMMENDATION SYSTEM

COLLABORATORS

1 RUTH KITASI
2 AGATHA NYAMBATI
3 JOSELINE APIYO
4 CECILIA NGUNJIRI
5 JOHN MBEGO
6 LEONARD KOYIO

OVERVIEW

Recommendation systems are essential for filtering information and improving user experiences.

Our project aims to develop an efficient movie recommendation system by analyzing users' past ratings to provide personalized movie suggestions, enhancing their viewing experience.

BUSINESS UNDERSTANDING: Navigating the Age of Data Overload with Recommendation Systems

The Challenge

- **Content Overload**: Users face overwhelming content choices on platforms like MovieLens.
- **Limited Discovery**: Traditional search methods are generic, leading to frustration and disengagement.

Our Solution

- **Recommendation Systems**: Analyze user interactions (ratings, behavior) to provide personalized movie suggestions, enhancing discovery and engagement.

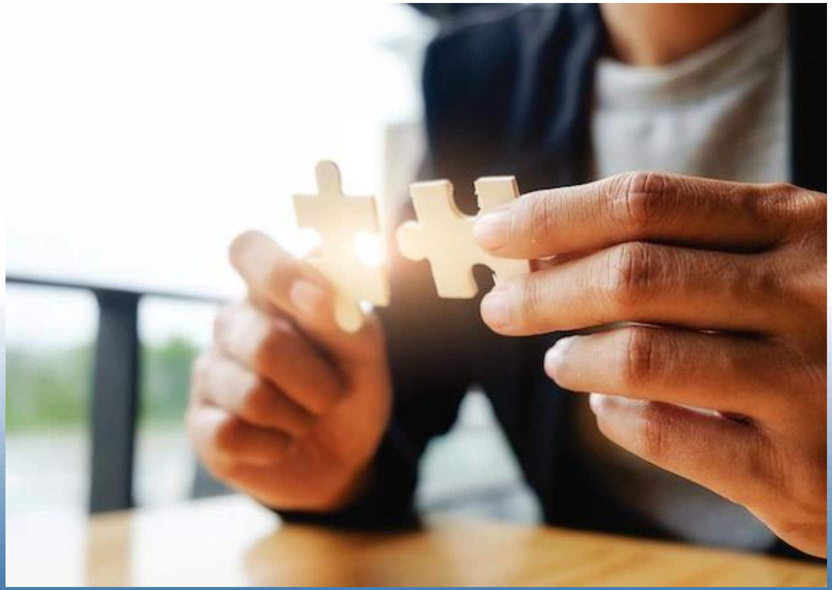
Business Impact

- Companies like Youtube and Spotify use similar systems to boost user engagement and retention by delivering curated content tailored to individual preferences.

PROBLEM STATEMENT

With the vast amount of content on streaming platforms, users often feel overwhelmed by choices, making it difficult to discover movies that align with their preferences.

Traditional search methods fail to address this challenge, resulting in a less satisfying user experience and decreased engagement.



DATA SOURCE

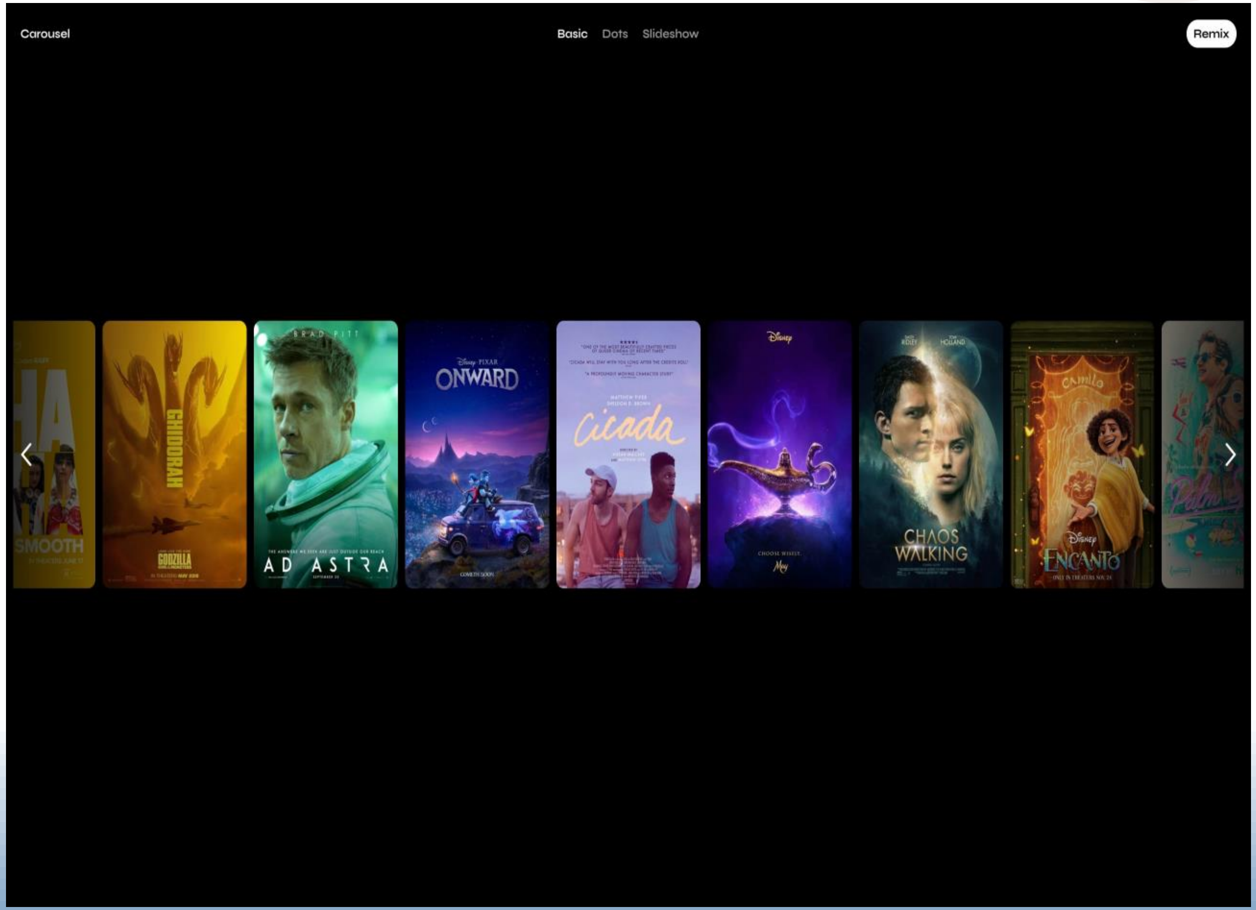
Origin: Sourced from GroupLens, a key resource for recommendation system research.

- **Personalized Recommendations**: Top 5 movie suggestions based on user ratings.
- **Content-Based Filtering**: Recommend similar movies based on user-input titles.
- **Cold Start Solutions**:
  - **Popular Movies**: Suggest top-rated films to new users.
  - **Genre-Based**: Recommend top films from selected genres for new users.
- **Performance Evaluation**: Use RMSE to evaluate recommendation accuracy.

DATA LIMITATION

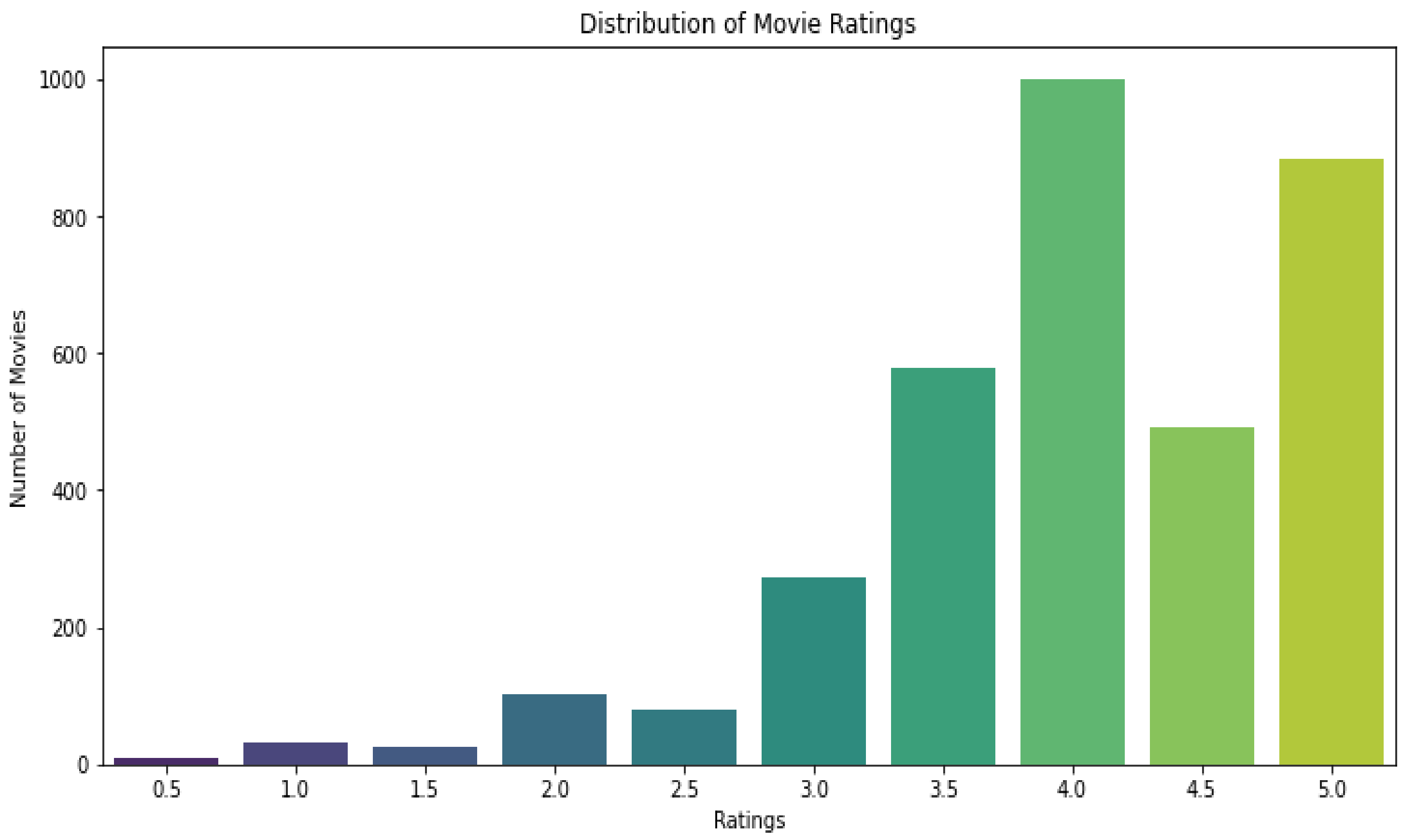
- Outdated Data: Limited recent movie information.
- Limited Genres: Dataset may not cover all movie genres.
- Cold Start Issue: Few ratings for new users and movies.
- Rating Bias: Popularity may influence user ratings.

IMPORTANT VARIABLES



- RATINGS
- TITLES
- GENRER
- USER ID
- YEAR OFPROUCTION

Movie Rating Distribution: How ratings are spread on a 0.5 to 5.0 scale.



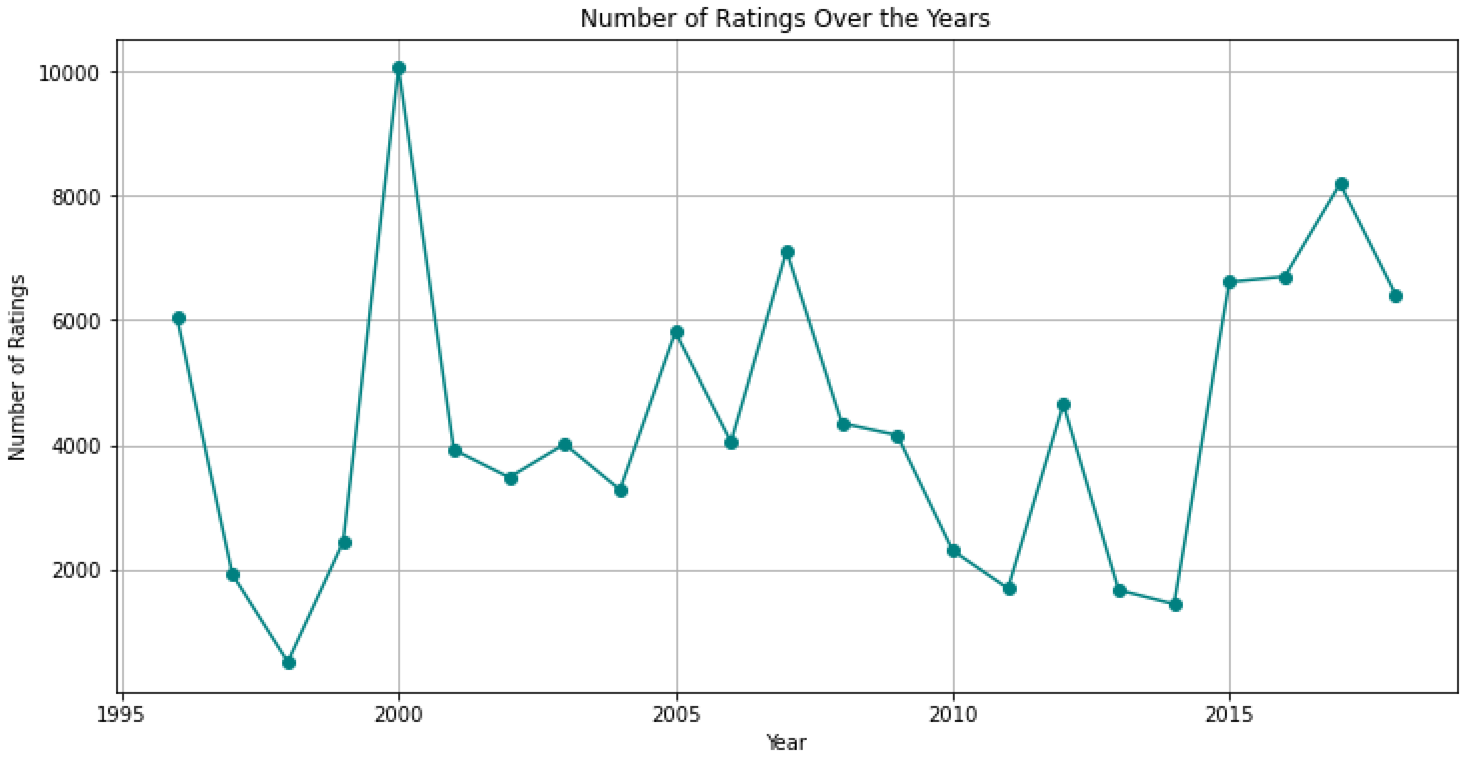
On a scale of 0.5 to 5.0, the analysis shows that most movies received an average rating of 4.0, indicating that users generally rated the majority of films positively.

This suggests a tendency for users to favorably evaluate the available content, with few movies receiving extremely low ratings of 0.5

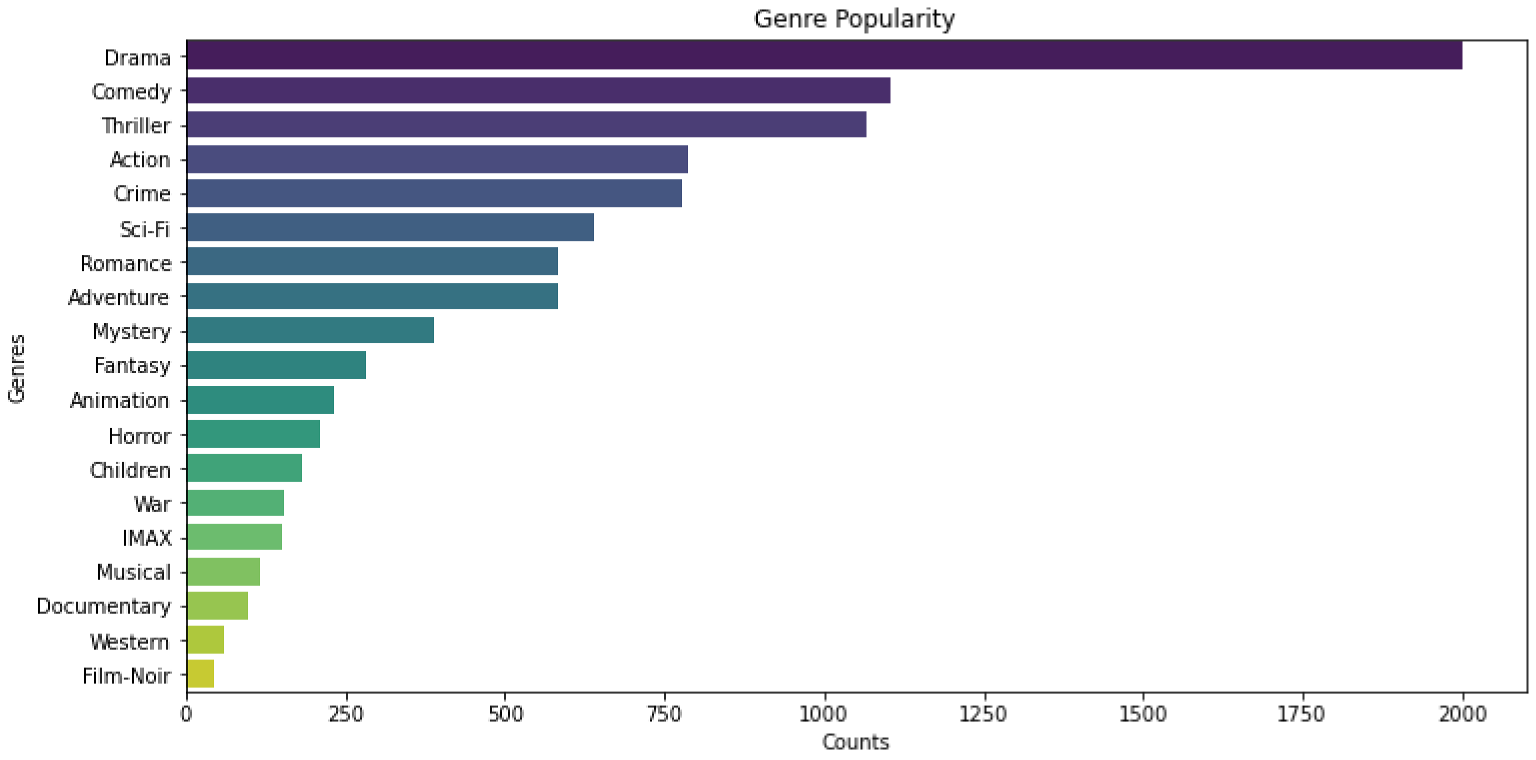
Rating Trends Over Time: Distribution of ratings across different years.

There is a noticeable spike around the year 2000, where the number of ratings peaked at over 10,000.

This implies that there was a surge in user activity, perhaps due to the popularity of certain movies or the increased availability of the platform at the time.



Genre Popularity: Which genres are most common in the dataset.

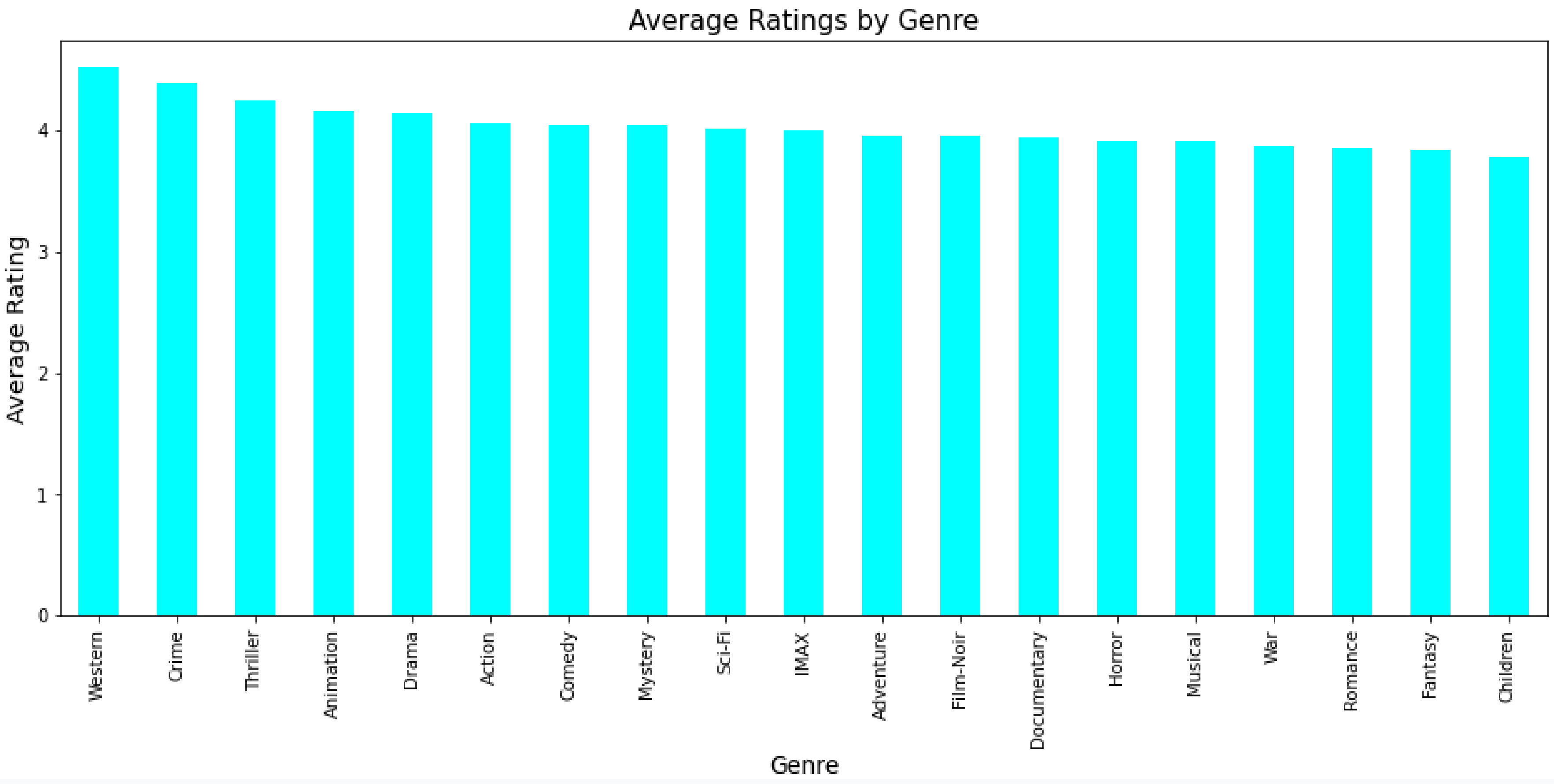


The graph shows that the genres "Drama" and "Comedy" are the most popular among the movies in the dataset, with significantly higher counts compared to other genres.

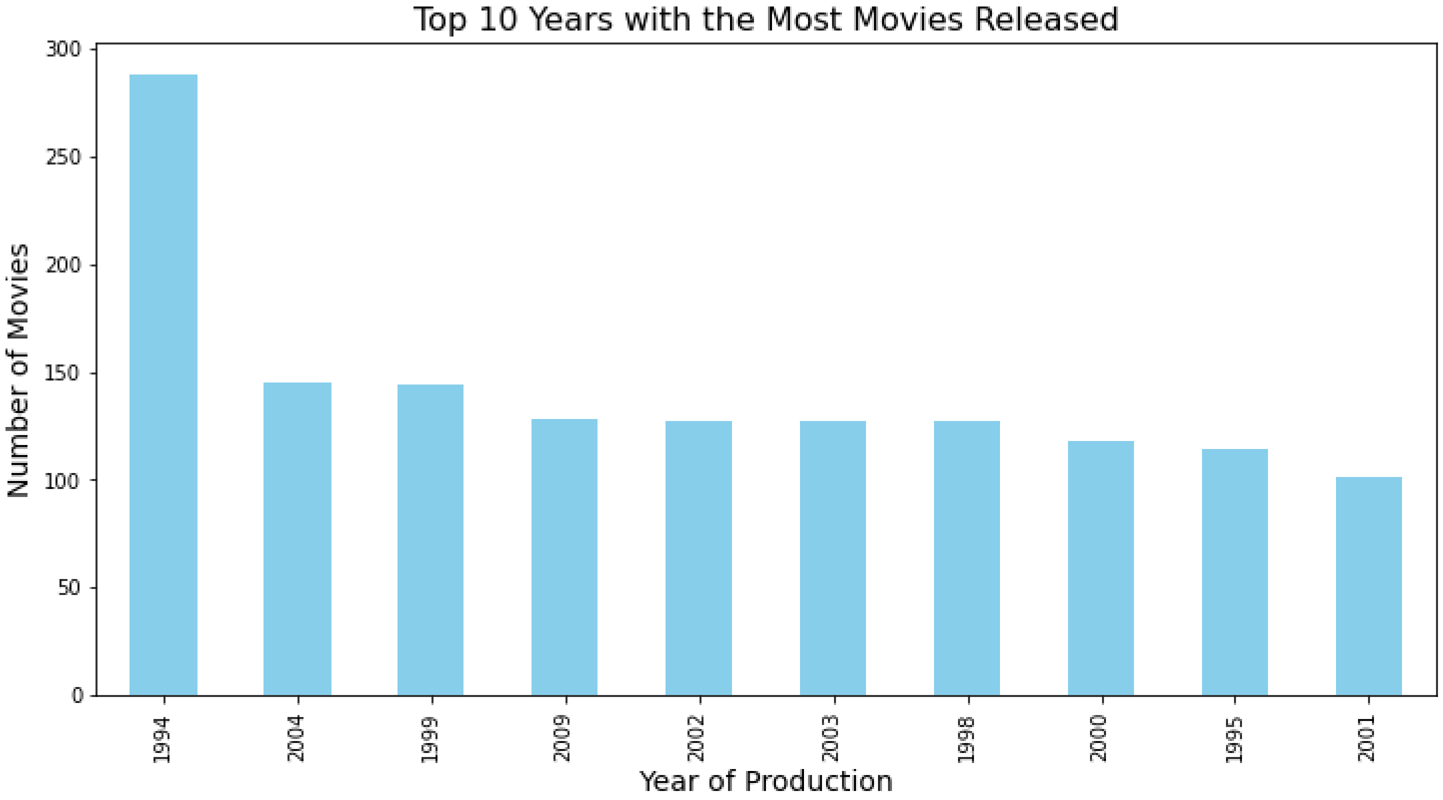
Average Ratings by Genre: Identifying top-rated genres.

Despite the relatively lower number of Western movies produced, this genre stands out due to its impressive average ratings, surpassing those of other genres.

This observation suggests that while Western films may not be as prolific as others, they resonate more strongly with audiences, gaining higher appreciation and positive feedback.



Release Years: Discovering the years with the highest number of movie releases.



The analysis indicates that the majority of films were produced in 1994.
One might expect this year to reflect the highest ratings. However, the ratings distribution graph reveals that 2000 had the highest number of ratings, despite a lower volume of films produced in that year compared to 1994.

MODELING

Building the Baseline Model:

● **Method**: KNNBasic (Surprise Library)
● **Approach**: Traditional collaborative filtering using cosine similarity.

● **Performance**:
○ **RMSE**: 0.7574
○ **Conclusion**: Higher RMSE indicates less accurate predictions.

● **Method**: Singular Value Decomposition (SVD)
● **Performance**:
○ **RMSE**: 0.53
○ **Conclusion**: SVD outperformed KNNBasic with significantly better accuracy.

Fine-Tuning the SVD Model

We focused on optimizing the SVD model to achieve the best predictive performance.
○ Performance:
■ Best RMSE: 0.50

2. Rounding Predictions:
○ Ensured predictions align with the dataset's 0.5-star increment rating scale.
Conclusion: The tuned SVD model with optimized hyperparameters delivered the best performance, minimizing prediction errors.

HYBRID MODEL: CONTENT-BASED FILTERING & COLD START MITIGATION

We Enhanced recommendations using content filtering while addressing cold start challenges.

Approach:
● **User-Based Filtering**:
○ Generated top 5 recommendations tailored to individual user ratings.

● **Content-Based Filtering**:
○ Suggested movies based on attributes (genres, tags) using TF-IDF and cosine similarity.
Cold Start Mitigation:

● **Content-Based Filtering for New Users**:
○ Recommended movies based on user-selected genres.
● **Movie Popularity**:

○ Suggested highest-rated movies irrespective of genre for new users.
Conclusion: The hybrid model successfully addressed cold start issues, ensuring high-quality recommendations for all users.

CONCLUSION

Personalized Recommendations: Collaborative filtering provided tailored suggestions.
Outcome: Boosted user engagement.

Content-Based Filtering: Users find similar movies by title.
Outcome: Enhanced personalization.

Cold Start Mitigation:
- **Popularity**: High-rated movies for new users.
- **Genre Selection**: Genre-specific recommendations for new users.

Outcome: Improved satisfaction for all user types.
Model Evaluation:
- **RMSE = 0.50**: Predictions within 0.50 points of actual ratings.

Outcome: Accurate recommendations.

Rating Frequency:
- **Average rating**: 4.0/5.0.

Outcome: Positive trends for future recommendations.

RECOMMENDATIONS

Recent Movies: Update dataset to include the latest releases for relevance.
Filmmaker Info: Add screenwriter and producer data for deeper personalization.
Broaden Genres: Expand to niche genres for wider audience appeal.
Location Data: Utilize user location for region-specific recommendations.
External Sources: Integrate reviews and social media sentiment to enhance quality.

Thank you