

Movie Recommendation System

The background features a dark teal gradient. Five 3D stars are arranged horizontally. The first three stars on the left are dark blue, the fourth is a lighter grey-blue, and the fifth is dark blue. A white, torn paper-like edge runs across the bottom of the stars, separating them from a solid black area at the very bottom.

Deon Durrant

Business and Data Understanding

Problem Statement

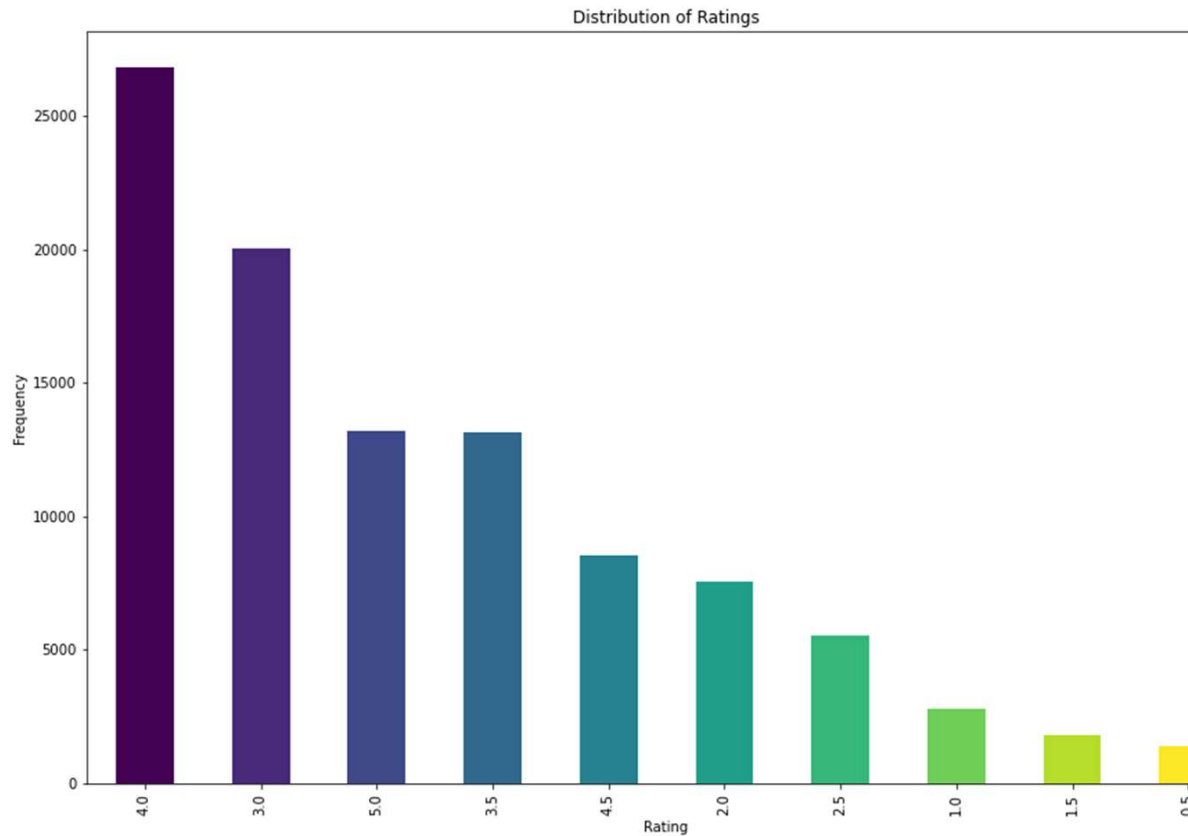
The King Isle event coordinator is responsible for organizing and recommending movie screenings for the residents of the island. The event coordinator needs to provide top five personalized movie recommendations to each resident based on their previous movie ratings. By analyzing the historical ratings given by residents to other movies, the system should predict which movies each resident would likely enjoy in the future and prioritize the highest-rated one

Business and Data Understanding

Rating Distribution

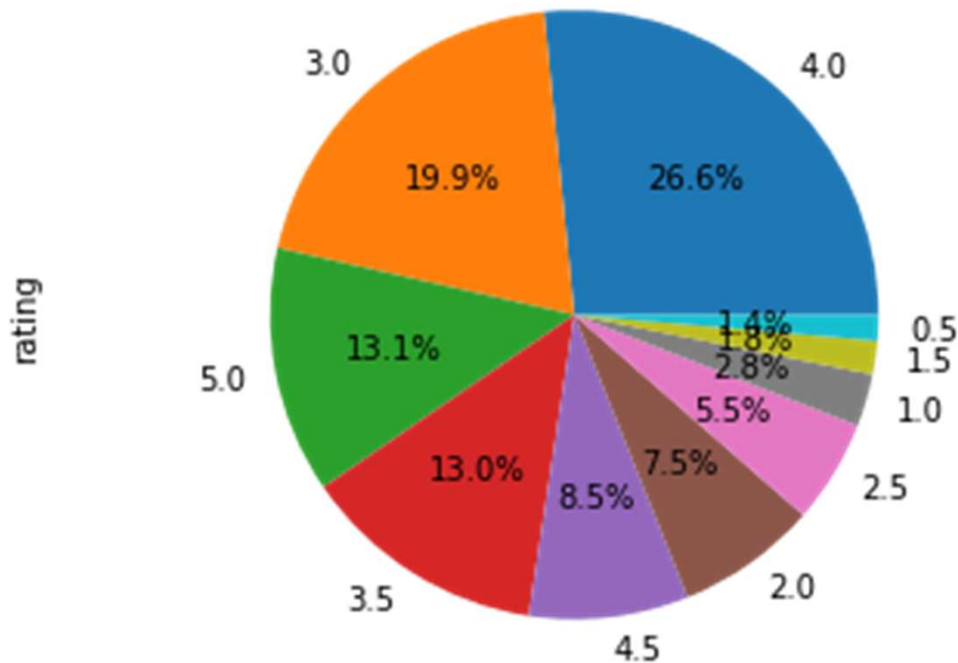
Rating are either implicit or explicit.

- Explicit rating scale **0-5**.
- Positive ratings of movies more common.
- Possible positive bias in the ratings..



Business and Data Understanding

Distribution of Ratings (Percentage)

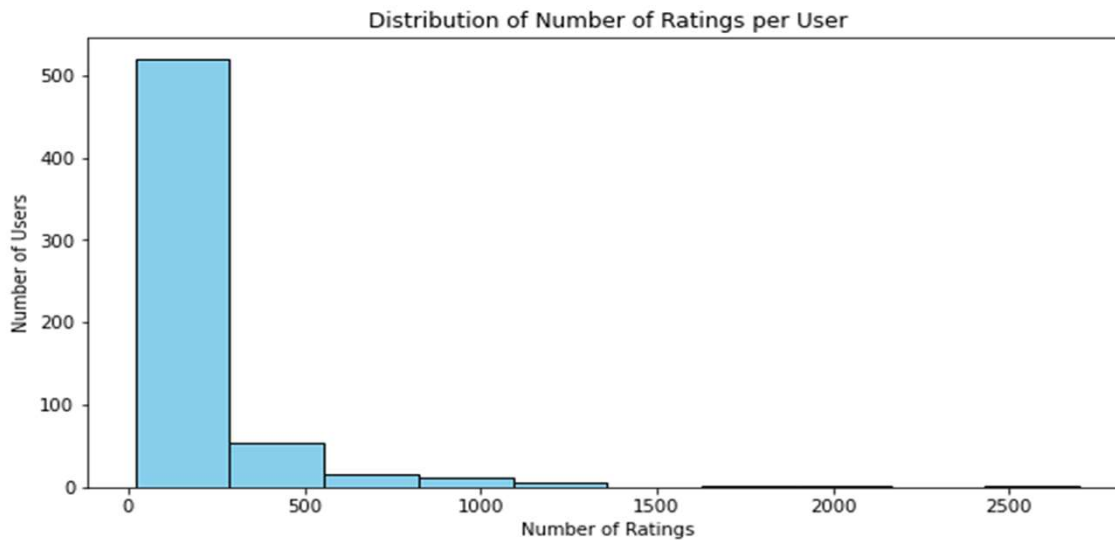


Ratings Percentage Distribution

Percentage distribution of ratings skews toward higher ratings

- **48.17%** ratings **4.0 -5.0**, yet **4.0 (26.60%)**
- **32.91%** ratings **3.0 to 3.5**
- **18.91%** of ratings below **3.0**

Business and Data Understanding



Distribution of ratings across users

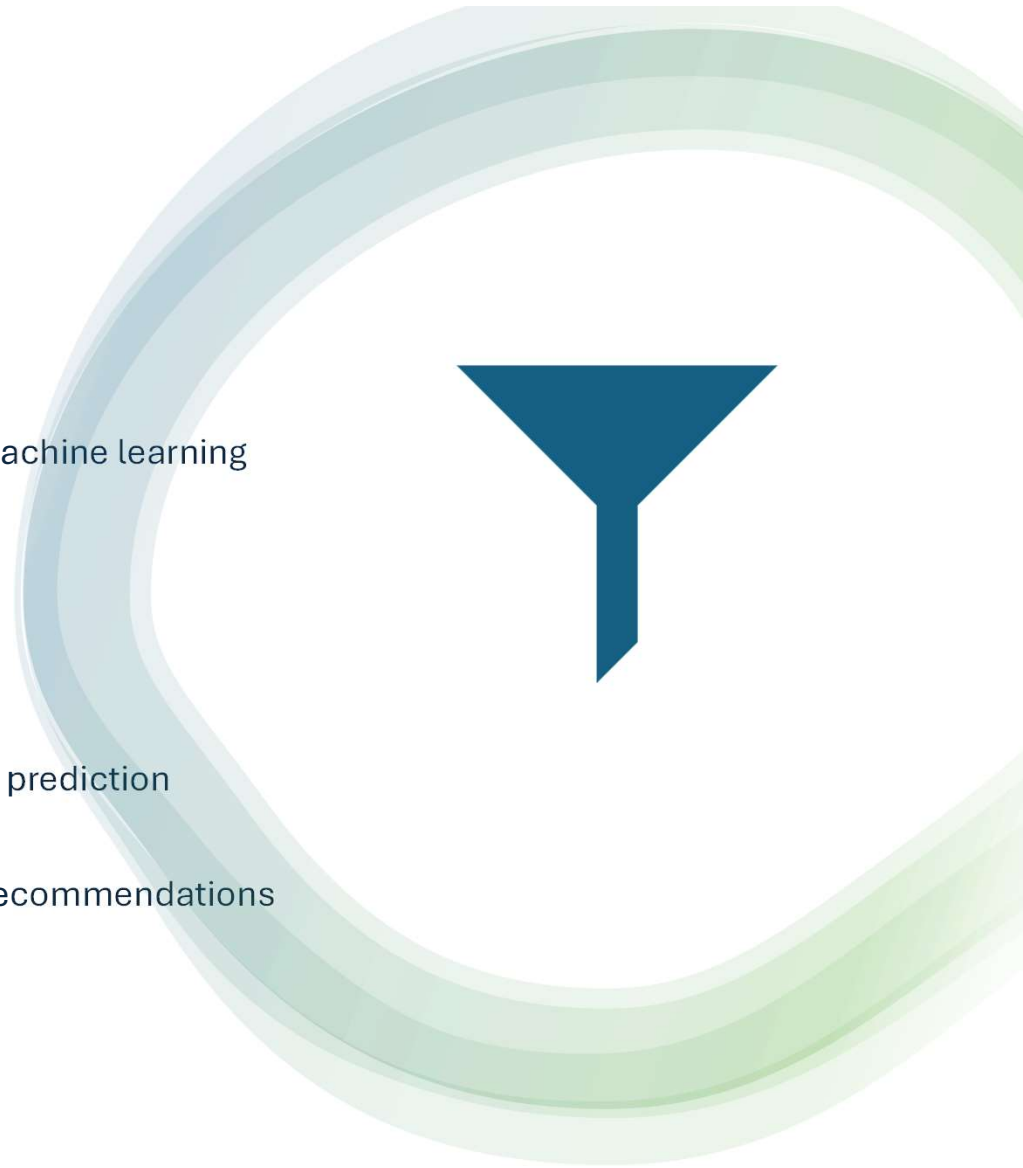
- Most users fall into the first bin (**approx. 250 ratings**)
- Low user engagement
- Suggest data sparsity
- **Fewer** users with **high rating** engagements

Recommendation System: Technique and Model selection

- Model-based Collaborative filter- creates a predictive machine learning model of the data
- Alternating Least Squares(ALS) Method

Evaluation Metrics

- Root mean square error RMSE- accuracy of the model's prediction
- Precision@K- percentage of relevant items in a list of recommendations



Recommendation System Objectives and Implementation

To address the problem, the recommendation system was designed to:

- Generate the top 5 recommended movies for each user
- Generate the top 5 recommended users for each movie

Proof of concept

- Recommend 5 movies to distinct users
 - Recommend 5 distinct movies to 5 user
-
- **Personalized Recommendations**



Top 5 Users Recommendation for Each Movie

userId	Title	Predicted Rating
53	Dracula: Dead and Loving It (1995)	3.548811
558	Dracula: Dead and Loving It (1995)	3.3245075
327	Dracula: Dead and Loving It (1995)	3.2381494
371	Dracula: Dead and Loving It (1995)	3.0953596
276	Dracula: Dead and Loving It (1995)	3.0835876

Performance Metrics for Movie Recommendation System

Metric	Top 5 Recommended Movies per User	Top 5 Users Recommendation for Each Movie
Root Mean Square Error (RMSE)	0.4464 Strong prediction accuracy with deviations of 0.44 units.	0.6255 Moderate accuracy with 0.63 units deviation from actual ratings.
Precision@K (4.5 relevance)	0.6614 66% of recommended movies meet threshold.	0.2830 28.3% of the top 5 recommended users for each movie met the threshold.
Precision@K (4.0 relevance)	0.8763 87% of recommended movies meet a lenient 4.0 threshold.	0.5944 59.4% of the top 5 recommended users for each movie are considered relevant .

User-Centric Recommendation Comparison Summary

Metric	Top 5 Recommended Movies per User	Recommending 5 Distinct Movies to 5 Users	Recommending 5 Movies to 20 Distinct Users
Root Mean Square Error (RMSE)	0.4464	0.5131 Moderate accuracy with 0.51 units from actual ratings.	0.2510 High accuracy with predictions close to actual ratings.
Precision@K (4.5 relevance)	0.6614 66% of movies are relevant to the user.	0.36 36% of the recommended movies meet a strict 4.5 threshold.	0.84 84% of recommended movies meet a strict 4.5 threshold.
Precision@K (4.0 relevance)	0.8763 87% of movies are relevant to the user.	1.0 All (100%) recommended movies meet a lenient 4.0 threshold.	0.96 96% of recommended movies meet a lenient 4.0 threshold.

Demonstration of Personalized Recommendations

userId	Title	Genres	Predicted Rating
54	Dragon Ball Z: The History of Trunks (Doragon bôru Z: Zetsubô e no hankô!! Nokosareta chô senshi - Gohan to Torankusu) (1993)	Action, Adventure, Animation	4.0924816
54	On the Beach (1959)	Drama	4.0924816
54	Seve (2014)	Documentary, Drama	4.00343
54	Deathgasm (2015)	Comedy, Horror	3.9777083
54	Saving Face (2004)	Comedy, Drama, Romance	3.9490638

Demonstration of Personalized Recommendations

Relevance Threshold	Precision@K	Interpretation
4.0	0.4	40% of the top 10 recommended movies are considered relevant by the user.
4.5	0.0	None of the top K recommended movies were considered relevant 0% .

Conclusion

The collaborative filtering system successfully addressed the business needs by offering relevant and high-quality recommendations.

Prediction accuracy RMSE range: **0.2510- 0.6255** indicating highly to moderately accurate predictions.

Precision@K:

- Better performance with a lenient threshold of **4.0**, 96% relevance in recommendations.
- Struggled with conservative **4.5** threshold

Effectively balances prediction accuracy and relevance



Next Steps

- Refine Model Parameters and Improve Accuracy
- Incorporate Hybrid Recommendations
- Expand Evaluation Metrics:

Mix of accuracy metrics and qualitative metrics

- Recommendations Refinement:

Feature engineer and incorporate implicit feedback to better capture user preferences and improve recommendation quality for individual users

Questions



Thank You!

Email: DurrantDeon@gmail.com

LinkedIn: [linkedin.com/in/dr-d-durrant](https://www.linkedin.com/in/dr-d-durrant)