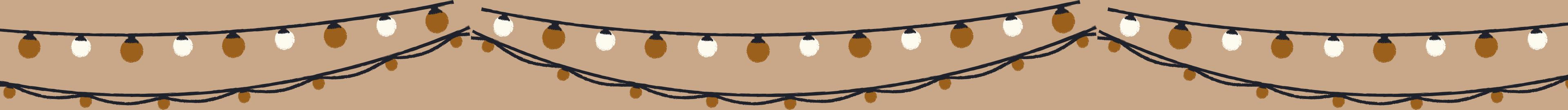




MOVIELENS RECOMMENDER SYSTEM



By Lorna Wanjiku



OVERVIEW



Introduction



Modelling



Business Understanding



Conclusion



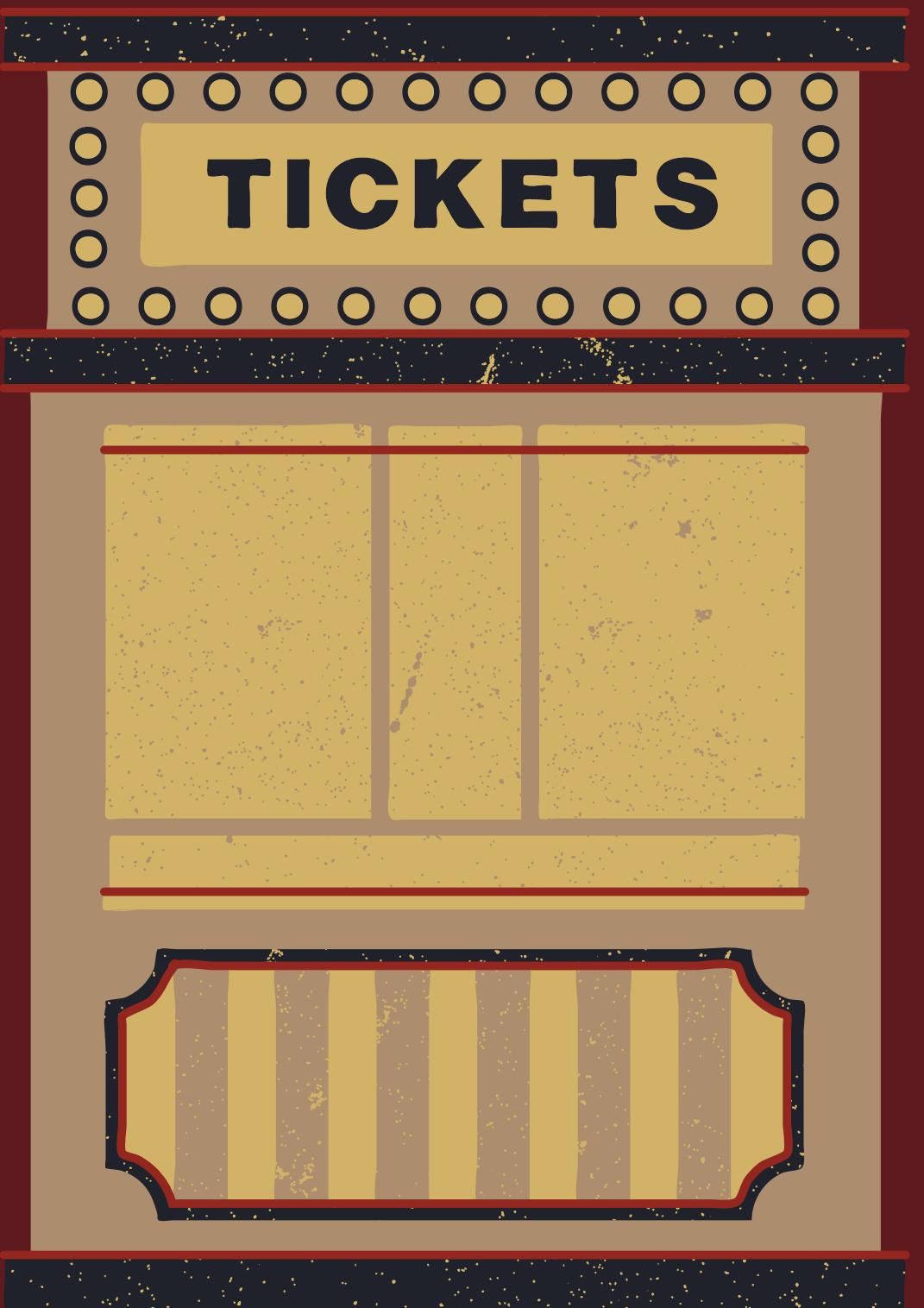
Problem statement



Recommendations

INTRODUCTION

- In today's vast selection of movies, finding ones that match user preferences can be challenging.
- Streaming platforms like Netflix address this issue by offering a wide variety of content. However, users often struggle to discover shows that align with their tastes due to the extensive library. Recommendation systems provide a solution by offering personalized movie recommendations based on user ratings. In this task, I will build a recommendation system using the MovieLens dataset to provide users with their top 5 movie recommendations.



BUSINESS UNDERSTANDING

- The objective is to build a recommendation system for Netflix to deliver seamless and personalized user experience by offering accurate movie and TV recommendations. As the Data Scientist involved my goal of this recommendation system is to enhance the movie-watching experience for users by suggesting movies that they are likely to enjoy based on their ratings of other movies. By offering personalized recommendations, one can increase user engagement, satisfaction, and ultimately drive user retention on the platform



STAKEHOLDERS

- Netflix Users - The primary stakeholders, since they rely on the recommendation system to enjoy relevant movies based on their preferences.
- Netflix Management and Executives - They want to improve user satisfaction, increase viewership and maximize revenue.



PROBLEM STATEMENT

- The problem is to build a recommendation system that can accurately predict and recommend the top 5 movies for a user based on their ratings of other movies and genres. The system should leverage collaborative filtering techniques and content-based filtering techniques to analyze user preferences and generate accurate recommendations.

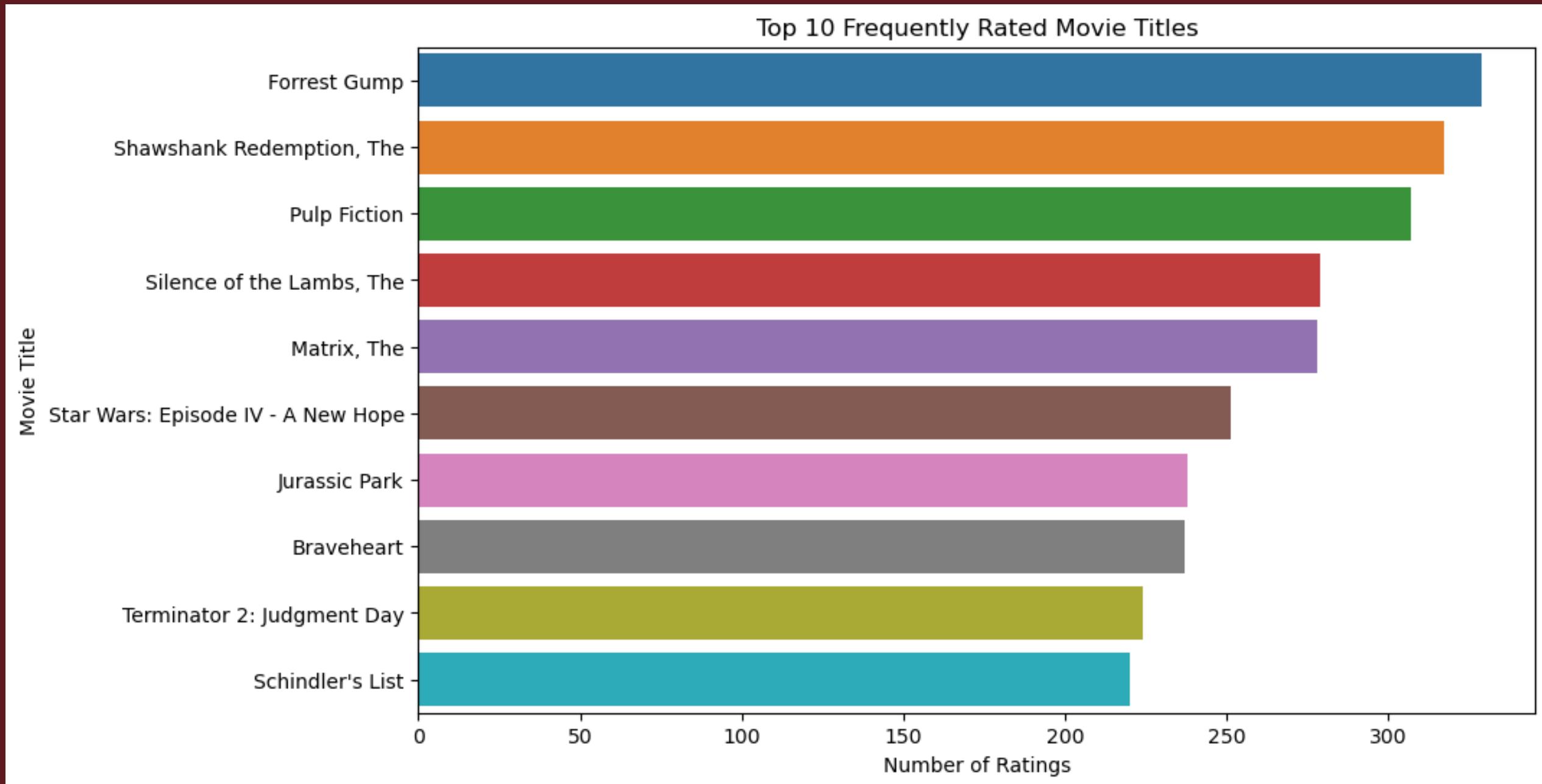


EXPLORATORY DATA ANALYSIS

Did Analysis to determine the relationships in the dataset.

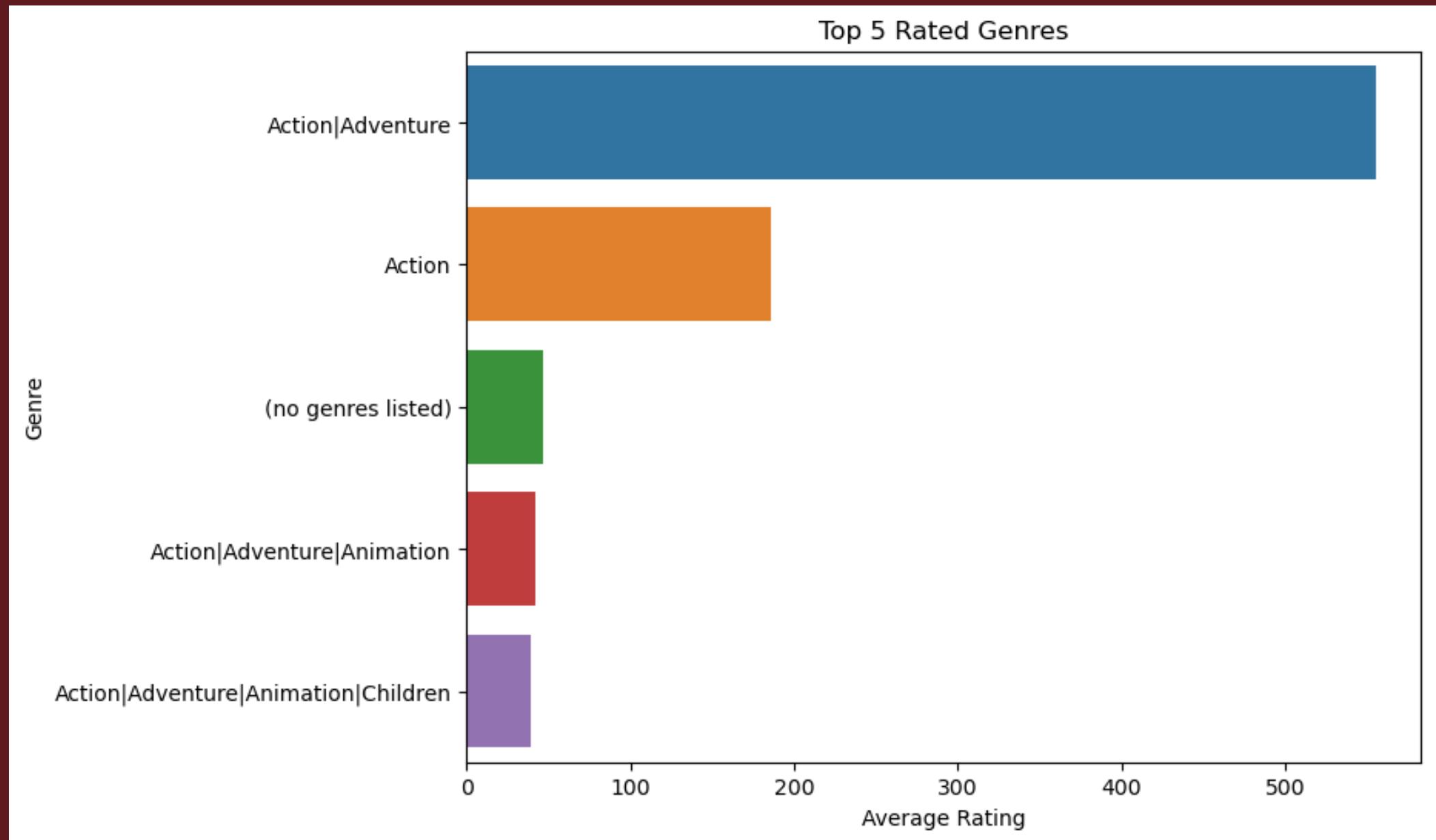


Frequently rated movie titles



Forrest Gump is the highest-rated movie(329)

Top 5 Rated genres



Action|Adventure is the highest rated genre

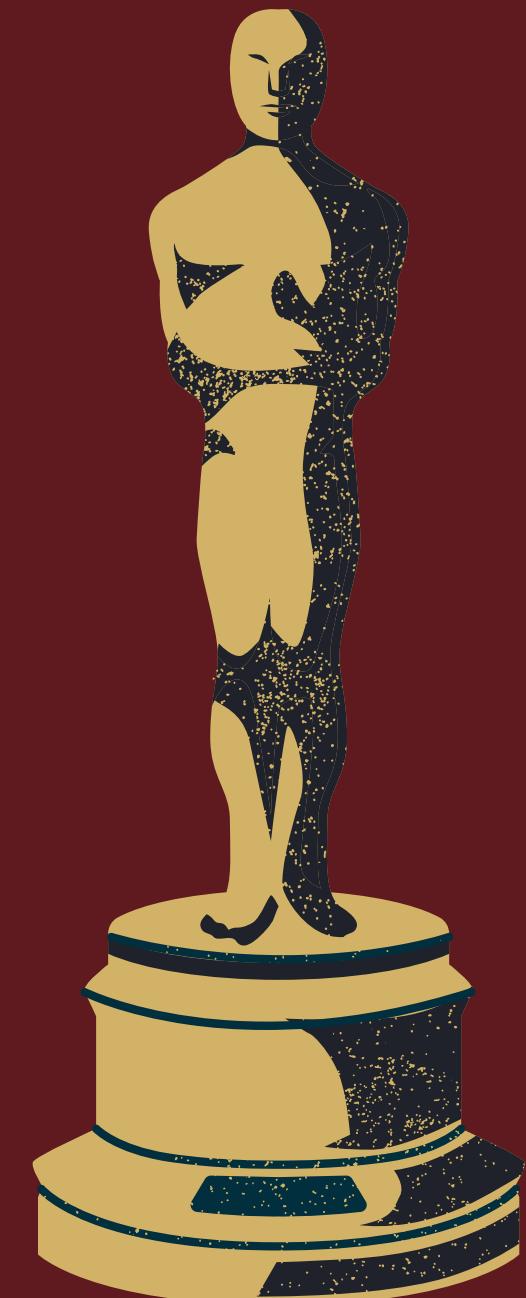
CONTENT-BASED FILTERING

Building a recommendation system that suggests movies based on their content features. In this context, movie genre as well as rating are used as content features to recommend similar movies.



MODELLING

- Matrix Factorization Methods like SVD(Singular Value Decomposition) and NMF(Non-Negative Matrix Factorization) were performed.
- MLP(Multi-Layer Perceptron)- Neural Networks of interconnected nodes.



Model 1

Matrix factorization
SVD - (Best Model)
Had the lowest RMSE(0.86)
and MAE(0.66)

Model 2

NMF
RMSE(0.92) and MAE(0.70) -
Above SVD values.

Model 3

KNNBaseline
RMSE(0.87) and MAE(0.67) -
Slightly above SVD values.

Model 4

MLP - Had the highest Values
might indicate overfitting.
RMSE(1.04) and MAE(0.83)

CONCLUSION

- EDA: - Action|Adventure is the highest rated genre.
- Content based filtering - Addressed the cold start problem where there is limited or no user data available for new users. Thus by analyzing content by genre and the ratings provided i managed to create a recommendation system based on similarities to the items.
- Modelling: - Matrix factorization(SVD baseline model and NMF) the SVD perfromed much better as compared to the other models (RMSE 0.85, MAE 0.66), the lower values indicate better accuracy and perfromance.
-



RECOMMENDATIONS

- Further analysis can be explored by feature selection and engineering and also perform some ensemble methods to improve on the metrics of success of the models.
- Getting more data on user feedback and user satisfaction also will aid in the improvement of the recommendation system.





THANK YOU

Don't hesitate to ask any questions!