

Movie Recommendation system

by Anita Makori !

Table of Contents

I n t r o d u c t i o n
P r o b l e m S t a t e m e n t
M a i n O b j e c t i v e
S p e c i f i c O b j e c t i v e s
N o t e b o o k S t r u c t u r e
D a t a U n d e r s t a n d i n g
M e t h o d o l o g y
C o n c l u s i o n s
R e c o m m e n d a t i o n s



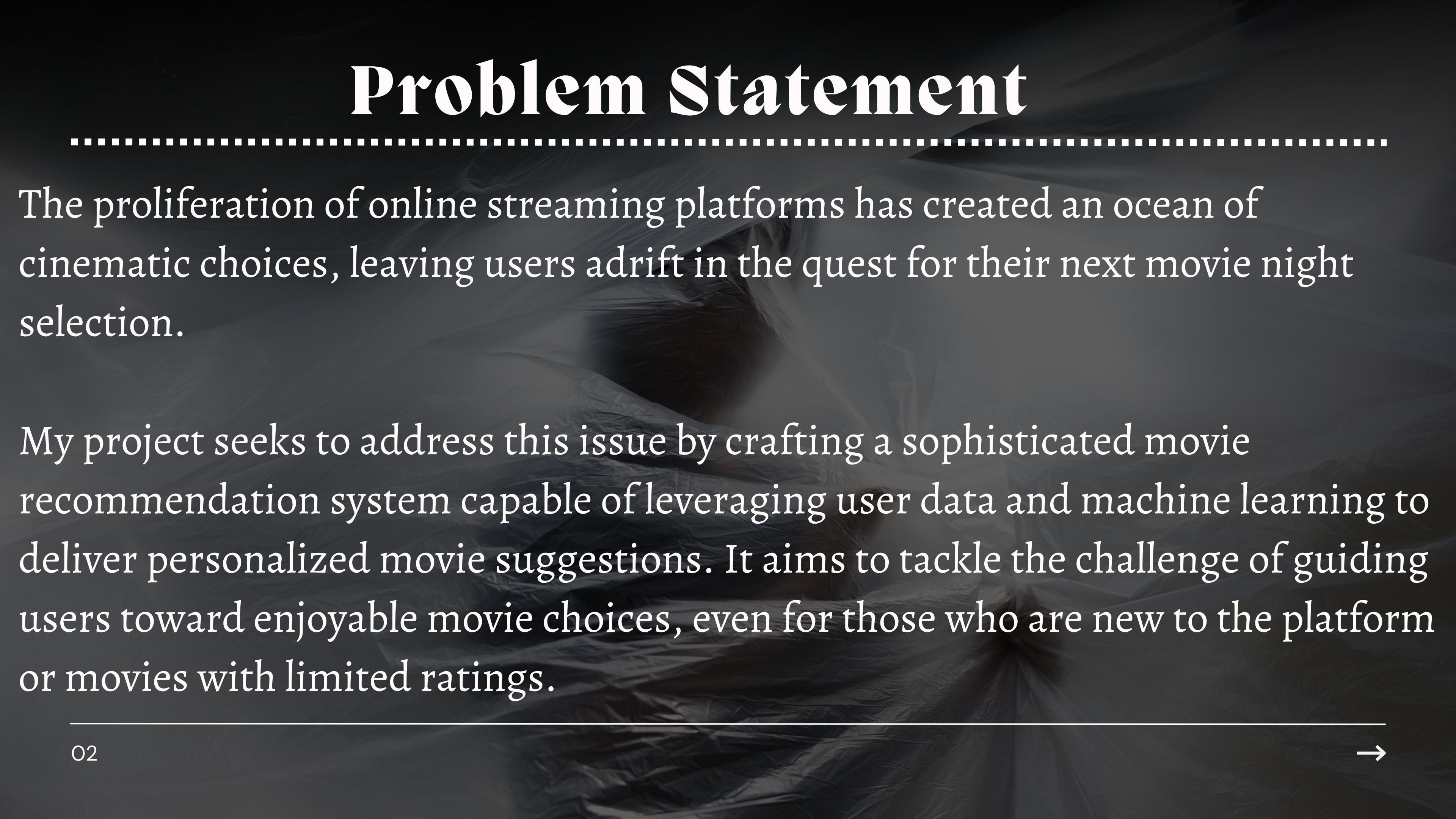
Introduction



In today's digital era, the world of cinema has witnessed a transformative shift. Streaming platforms have unlocked a vast library of movies, granting users unprecedented access to a multitude of entertainment options.

However, amidst this abundance, users often find themselves overwhelmed, struggling to identify movies that resonate with their individual tastes. In response to this challenge, I introduce an innovative movie recommendation system, poised to revolutionize the way users discover and enjoy films.

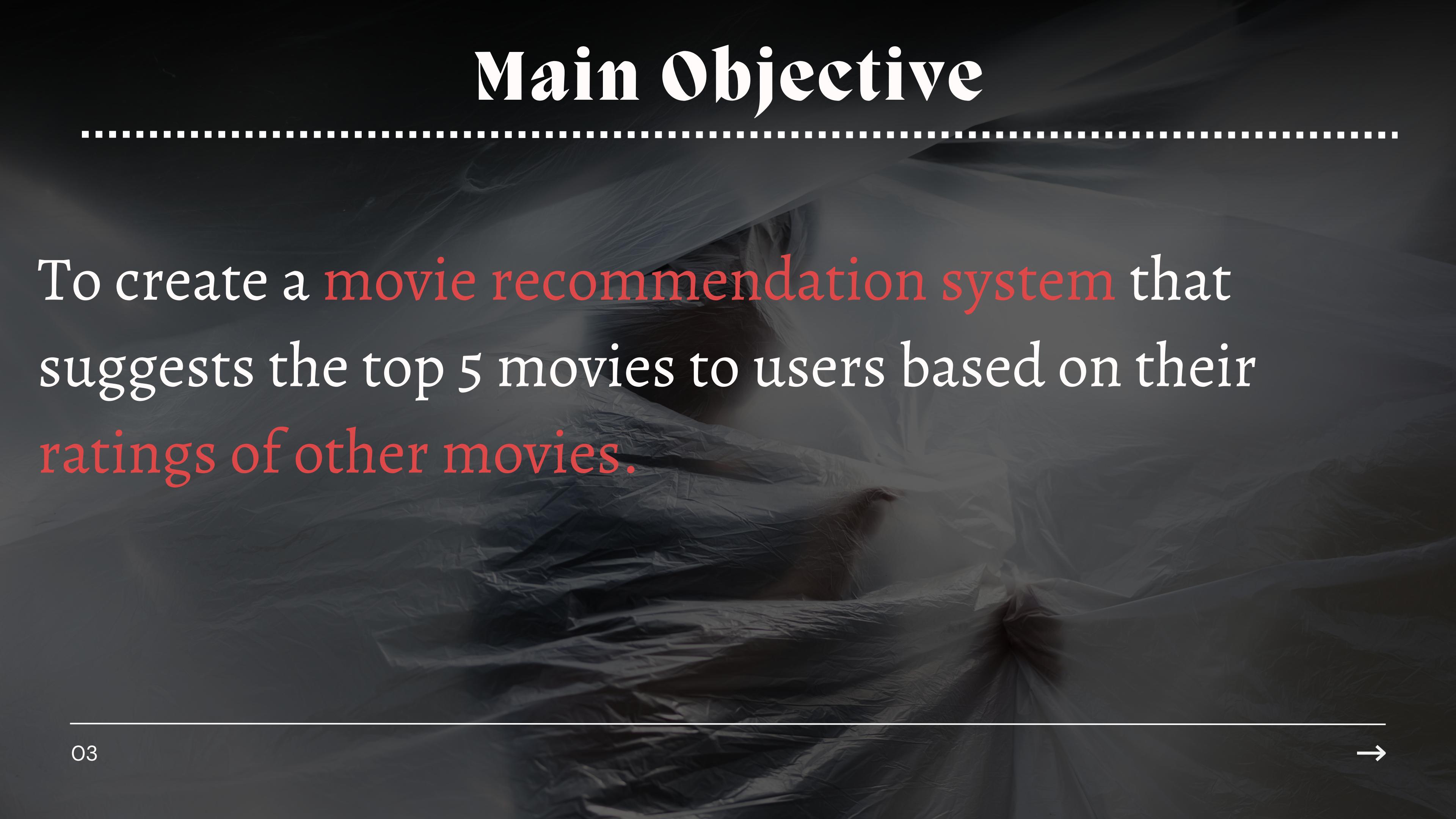
Problem Statement



The proliferation of online streaming platforms has created an ocean of cinematic choices, leaving users adrift in the quest for their next movie night selection.

My project seeks to address this issue by crafting a sophisticated movie recommendation system capable of leveraging user data and machine learning to deliver personalized movie suggestions. It aims to tackle the challenge of guiding users toward enjoyable movie choices, even for those who are new to the platform or movies with limited ratings.

Main Objective



To create a **movie recommendation system** that
suggests the top 5 movies to users based on their
ratings of other movies.

Specific Objectives



- To evaluate a comprehensive evaluation of user-based and item-based recommendation algorithm as well as a deep neural collaborative filtering
- To assess the overall user experience and engagement with the recommendation platform.
- To investigate the recommendation system's performance and scalability to accommodate a growing user base and movie catalog.

Notebook Structure

- Loading the Data
- Data Preprocessing
- Exploratory Data Analysis
- Modeling
- Results
- Conclusions
- Recommendations

Data Understanding

.....

1. Data Source

-The dataset used in this project for building a recommendation system was sourced from the GroupLens research lab at the University of Minnesota.

2. Data Features

-Movie Data: Information about movies is available, including movie IDs, titles, genres, release years, and potentially other metadata.

-Ratings Data: The core of the dataset consists of user ratings for movies, with each record typically containing user IDs, movie IDs, and the corresponding ratings.

Data Understanding

.....

3. Data Quality

- Missing Values: An initial check reveals that there are no missing values in the user IDs, movie IDs, or ratings fields.
- Data Consistency: Data consistency checks have been performed, and there are no apparent data inconsistencies or format issues.
- Outliers: An initial examination of ratings data suggests that there are no extreme outliers or unrealistic values.

Data Understanding

.....

4. Data Exploration

- Distribution of Ratings: The distribution of user ratings has been visualized, showing the distribution of ratings across the entire dataset. This information can help in understanding user preferences.\

- Movie Genre Distribution: A breakdown of movie genres and their frequency in the dataset has been examined. This helps in understanding the diversity of movie genres available.

Data Understanding

.....

5. Data Challenges

- Sparsity: The dataset may suffer from sparsity, where most users have rated only a small subset of movies. This can impact recommendation accuracy.
- Cold Start Problem: Handling new users or movies with very few ratings is a potential challenge. Strategies to address this issue need to be developed.

Data Understanding

.....

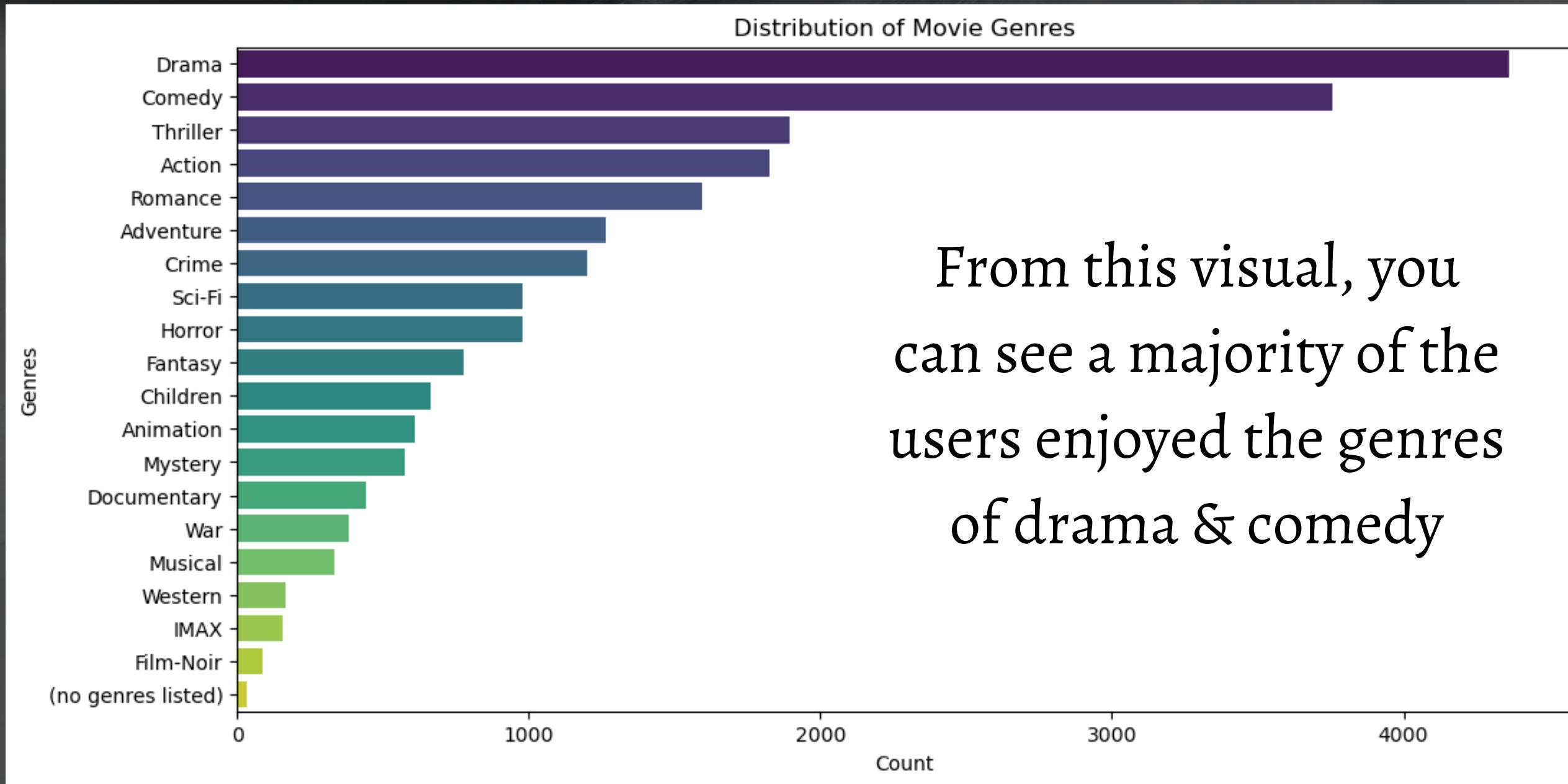
6. Data Preprocessing

-Data preprocessing steps such as dropping unnecessary columns, merging datasets will be applied as needed.

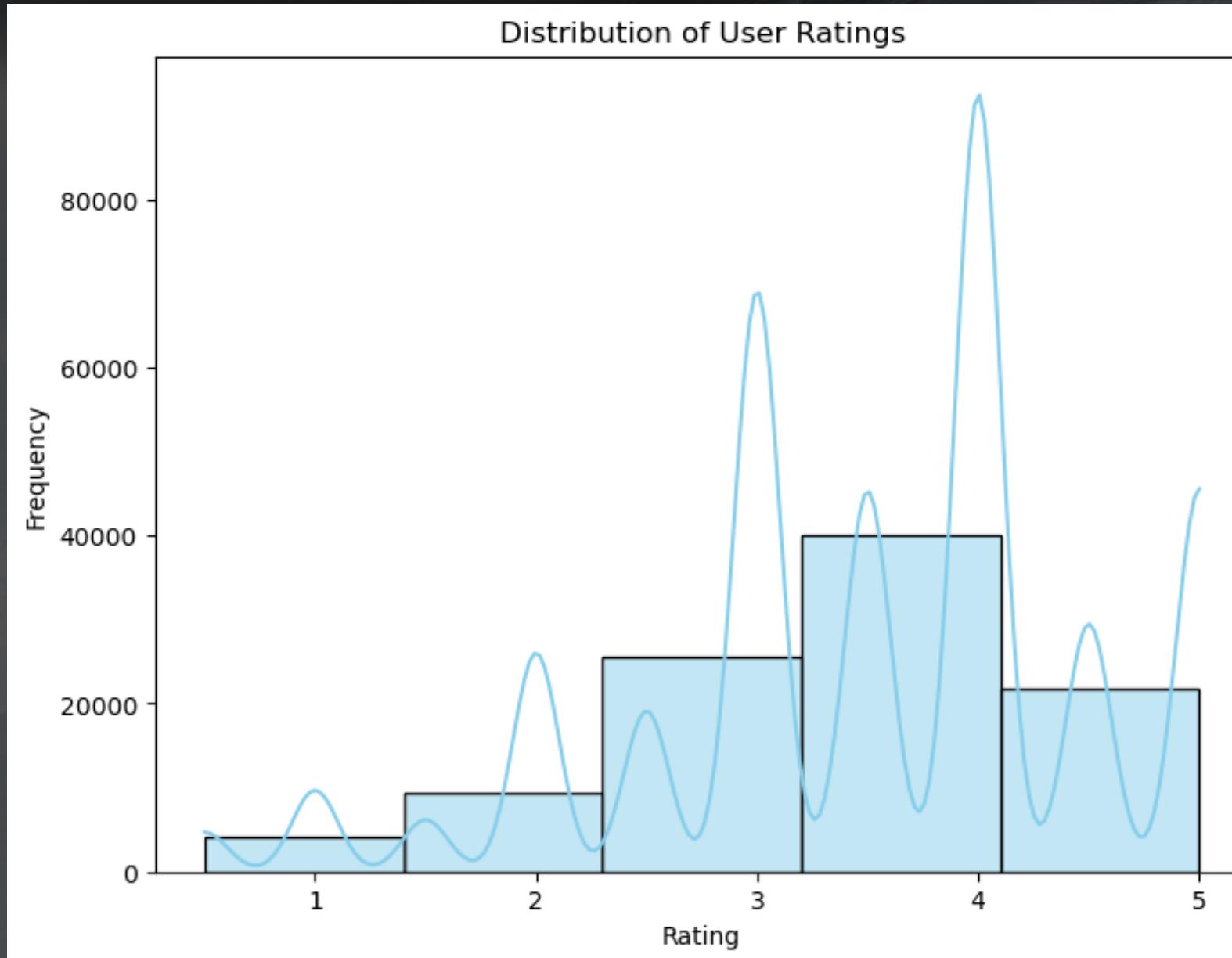
7. Data Splitting

-The data will be divided into training and testing sets for model evaluation. Standard cross-validation techniques may also be employed.

EXPLORATORY DATA ANALYSIS



EXPLORATORY DATA ANALYSIS



From the visualization, major users have rated movies from a range of 3-5

MODELING & RESULTS

.....

1. ****Start****: The model begins with relatively good predictions on movies users have seen during training (low training loss) and less accurate predictions on unseen movies (higher validation loss).
2. ****Training vs. Validation****: The model tries to improve throughout training. However, it's more accurate on the training data than on the validation data, which is normal.
3. ****End of Training****: After 10 rounds of training (epochs), the model's predictions on unseen movies have a validation loss of around 0.1005. This shows that the model can make decent recommendations but has room for improvement.

EVALUATION

.....

An RMSE of approximately 0.317 means that, on average, the model's predicted ratings for movies are off by around 0.317 rating units compared to the actual ratings provided by users.

An RMSE of 0.317 on the validation data indicates that the model is providing movie recommendations with reasonably good accuracy.

CONCLUSIONS

.....

1. Recommendation System Effectiveness: The project has successfully developed and implemented a movie recommendation system that leverages collaborative filtering techniques, including user-based and item-based approaches, as well as neural collaborative filtering (NCF). The system provides personalized movie recommendations to users based on their historical ratings and preferences.

CONCLUSIONS

.....

2. Model Performance and Validation: The project evaluated the performance of the recommendation models using metrics such as RMSE (Root Mean Squared Error) on a validation dataset. The RMSE values obtained for the different models indicate their predictive accuracy. The NCF model with an RMSE of approximately 0.316 has demonstrated better performance compared to other models, suggesting that it is more effective in predicting user preferences.

CONCLUSIONS

.....

3. User Engagement and Content Discovery: The recommendation system addresses the challenge of content discovery on streaming platforms.

By offering tailored movie suggestions, it enhances user engagement, reduces decision fatigue, and improves the overall user experience. Users can discover movies aligned with their tastes, even if they are new to the platform or if the movies have limited ratings.

RECOMMENDATIONS

.....

1. Continuous Model Refinement: To further enhance the movie recommendation system, it is recommended to continuously refine and optimize the recommendation models. This can include exploring advanced deep learning techniques, incorporating additional user and movie features, and experimenting with different architectures. Regular model updates based on user feedback and evolving preferences can lead to even more accurate and personalized recommendations.

RECOMMENDATIONS

.....

2. Diverse Recommendation Strategies: While the project primarily focused on collaborative filtering and neural collaborative filtering, it is advisable to diversify recommendation strategies. Incorporating content-based filtering, where recommendations are based on movie characteristics (genres, actors, directors), can provide a more comprehensive recommendation experience. Hybrid recommendation systems that combine multiple approaches can potentially offer a richer and more diverse set of movie suggestions.

RECOMMENDATIONS

.....

3. User Feedback and Interactivity: Implementing a feedback mechanism within the recommendation system can be valuable. Users should be encouraged to provide explicit feedback on recommended movies (e.g., thumbs up/down, ratings, comments). This feedback can be used to further refine recommendations and adapt to changing user preferences over time. Additionally, enabling users to explore and customize their recommendation criteria (e.g., filtering by genre, release year) can empower them to have more control over their movie discovery experience.