MOVIE

RECOMMENDATION SYSTEM

Phase 4, Group 12.

Movie Reco.

OVERVIEW

1. Introduction.
2. Objectives.
3. Data understanding.
4. EDA $Visualization.
5. Modelling.
6. Conclusion.
7. Recommendation.
8. Next steps.

Movie Reco.

INTRODUCTION

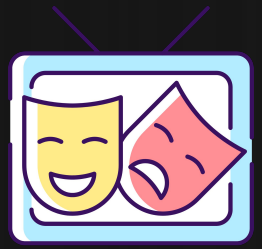
In today's digital age, content recommendation systems play a crucial role in enhancing user experience by providing personalized suggestions tailored to individual preferences. This project aims to build a recommendation model that offers users their top 5 movie recommendations based on their preferred genres and ratings of other films .

Movie Reco.

OBJECTIVES

Movies

To analyze and find the most frequently watched movies



Genres

To analyse and find the most frequently watched genres

Ratings

Create a system that makes movie recommendation based on user ratings by suggesting products that resonate with user's past

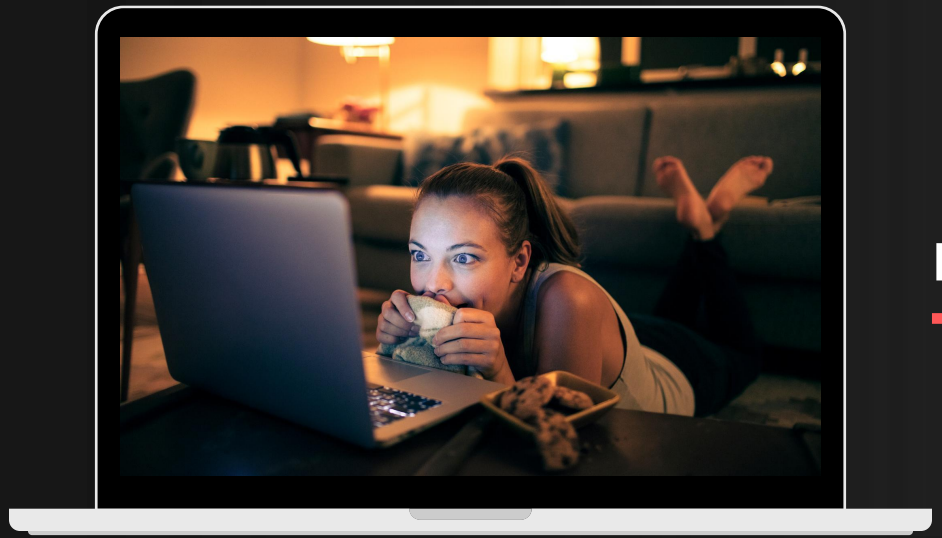
Movie Reco.

Data Understanding.

For this project we used MovieLens dataset from GroupLens research focusing on the movies and ratings dataset (100836 rows * 6 columns):

- **MovieId -** unique identifier of the movie
- **Title -** name of the movie
- **Genres -** athematic category of movie
- **UserId -** unique customer identifier.
- **Rating -** a measurement of the quality
- **Timestamp -** a digital record of the time of occurrence of a particular event.

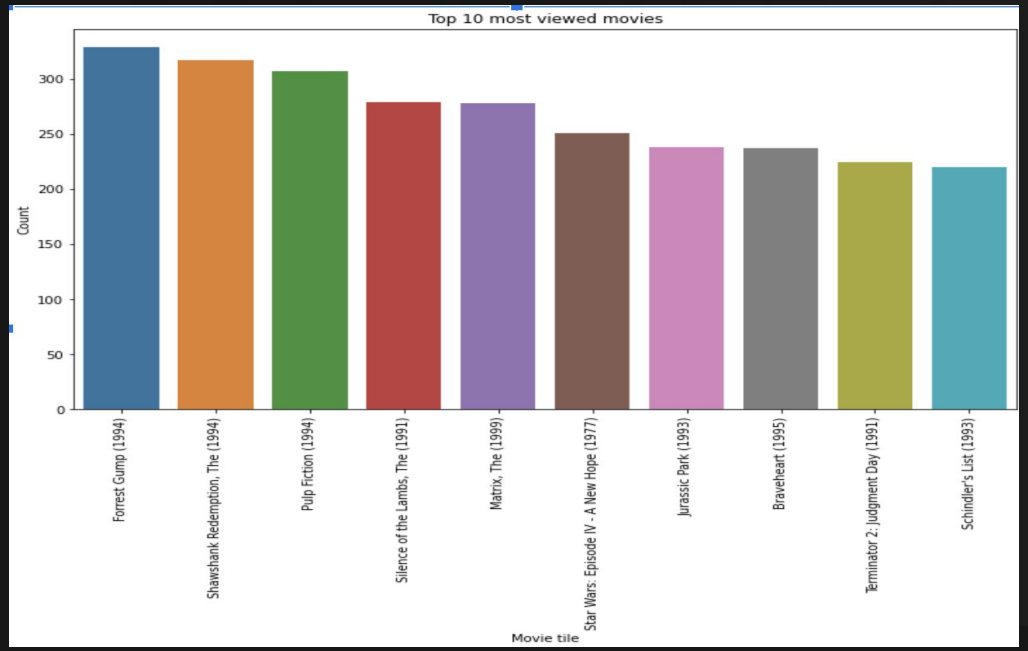
Movie Reco.



EDA & VISUALIZATION.

Movie Reco.

UNIVARIATE ANALYSIS

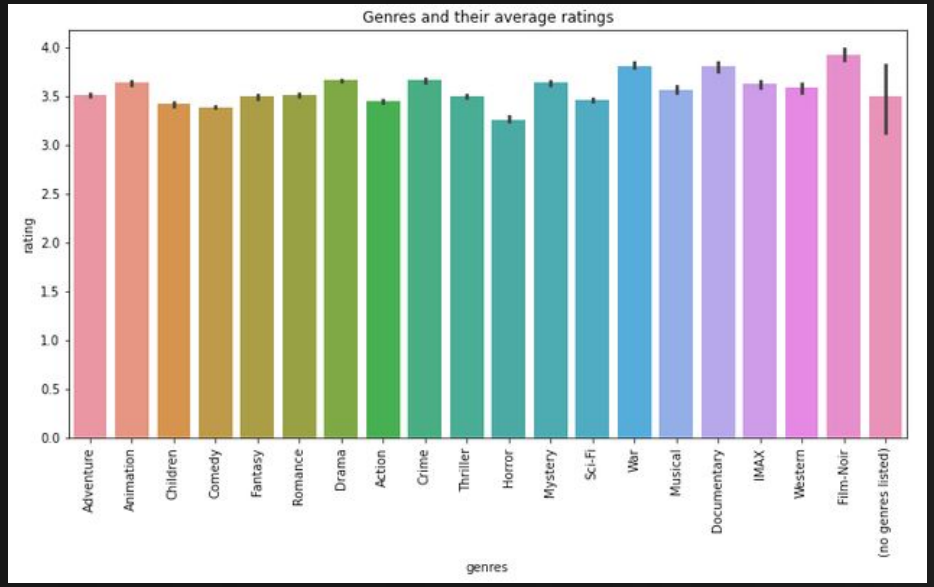


The most frequently viewed movies are Forrest Gump (1994) and Shawshank Redemption (1994) respectively..

The top 10 includes a mix of genres, including drama, comedy, action, thriller, and science fiction.

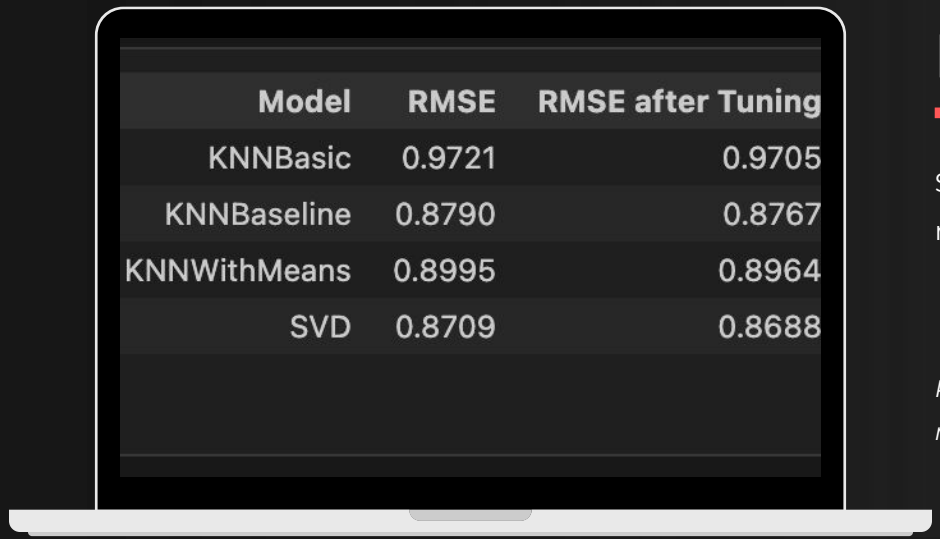
Movie Reco.

BIVARIATE ANALYSIS



From the bar chart it can be seen that all the genres have an average rating of above 3.

The ratings for the genres is higher in film noir which is stylized Hollywood crime dramas, particularly those that emphasize cynical attitudes and motivations followed by Documentary.



Model	RMSE	RMSE after Tuning
KNNBasic	0.9721	0.9705
KNNBaseline	0.8790	0.8767
KNNWithMeans	0.8995	0.8964
SVD	0.8709	0.8688

Modelling

SVD model exhibits superior prediction accuracy, recommended for deployment.

RMSE (Root Mean Square Error) reduction was used as a measure of predictive accuracy.

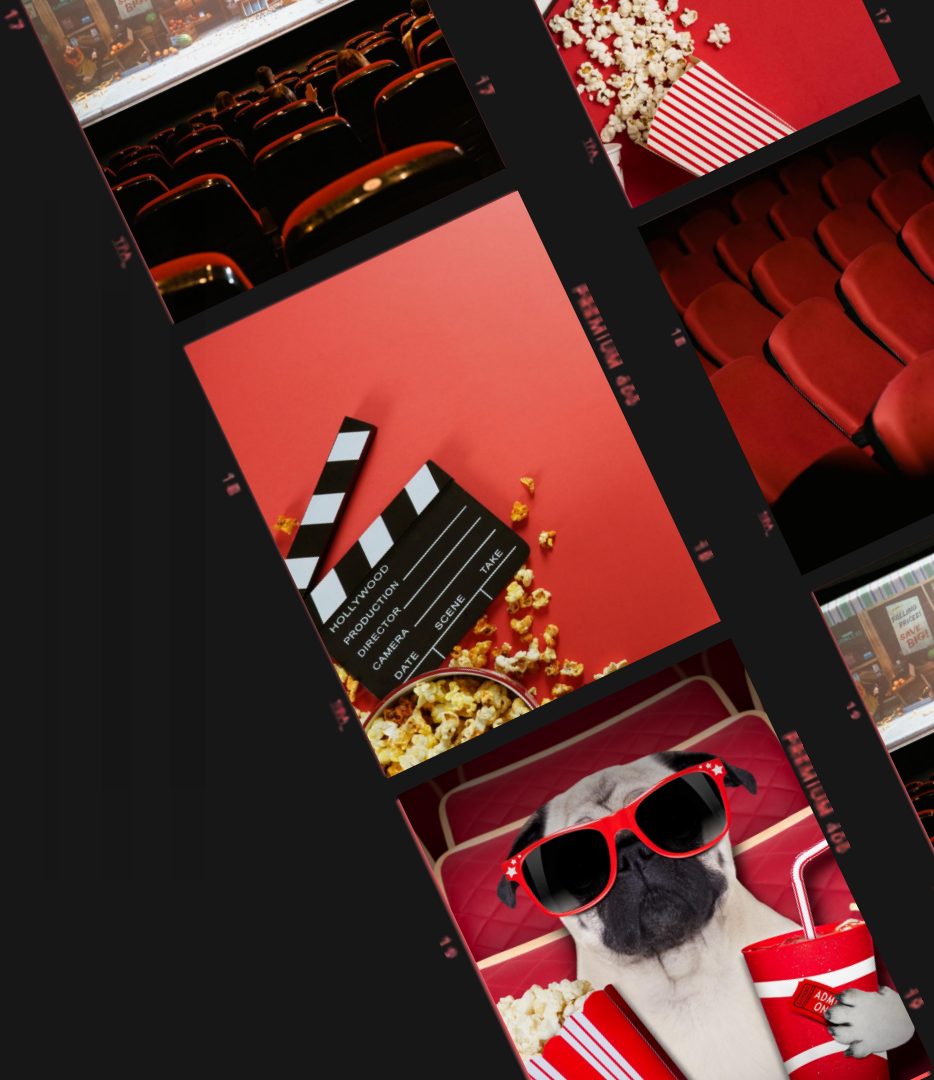
Movie Reco.

Deployment

Movie Reco.

Conclusion

- The top 3 movies are:
  -Forrest Gump(1994),The Shawshank Redemption(1994) and Pulp Fiction(1994)
- The bar plot shows the distribution of all the movie genres in the dataset.
- The 3 top genres are listed below starting from the most common one:Drama,Comedy and Action



Movie Reco.

Recommendation

Filamu should show these top movies: Forrest Gump, Hoop Dreams, and Pulp Fiction (all from 1994).

For popular genres, focus on Drama, Comedy, Action, and Thriller.

The recommendation system will help suggest enjoyable movies based on user preferences.

Next Steps

- Enhance scalability by implementing distributed computing frameworks like Spark to handle large datasets more efficiently.
- Create more comprehensive user and item profiles by incorporating features such as descriptions, demographics, and historical behavior.
- Improve user trust and engagement by implementing mechanisms to provide explanations for the recommended items.

THANKS

ANY QUESTIONS ??

Project by:

Mwangi Wambugu
Esther Nyawera
Grace Mutuku
Rony Mureithi
Heri Kimotho
Peter Otieno
John Kioko