



JUJU STREAMING SERVICE SOLUTION



METRICULAS
DATA CONSULTANTS



MELANIA MAKOMBE
CEO



ABUBAKAR SADIQ
DATA ANALYST



MICHAEL KANU
API, GITHUB
SPECIALIST



NURADDEEN ISAH
JUNIOR DATA
ENGINEER



**CHRISTIAN
DIVINEFAVOUR**
SENIOR MODELING
ENGINEER

ABOUT US

- Operating since 1995
- Major clients include Tesla and Netflix
- Website: www.metriculas.co.za

PRESERNTATION OUTLINE

INTRODUCTION

Recommender system overview

PROBLEM STATEMENT

Understanding the problem

DATA PREPROCESSING

Cleaning and preprocessing the data

EXPLORATORY DATA ANALYSIS

Understanding the data morphology

MODELING

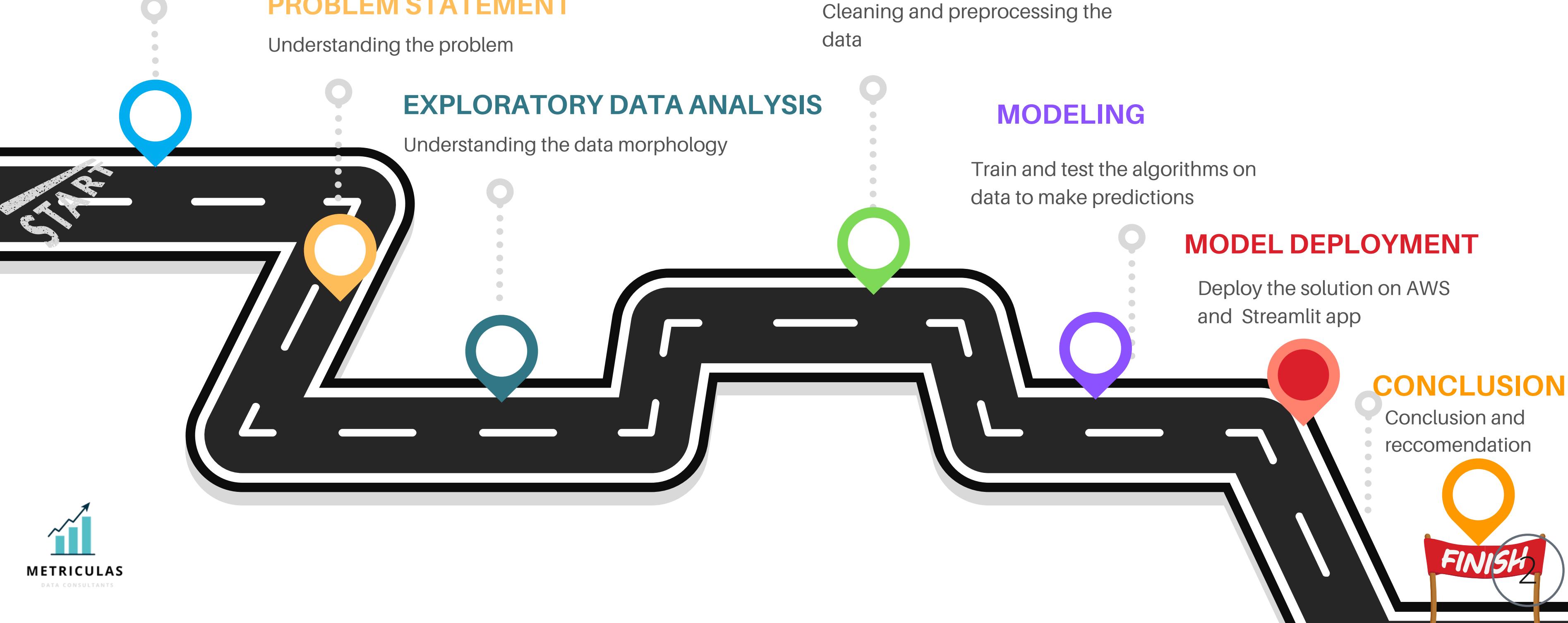
Train and test the algorithms on data to make predictions

MODEL DEPLOYMENT

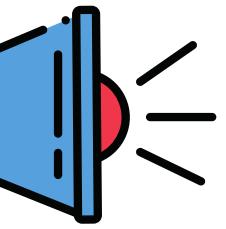
Deploy the solution on AWS and Streamlit app

CONCLUSION

Conclusion and recommendation



INTRODUCTION



**69% CONSUMERS PREFER VIDEO STREAMING
OVER TRADITIONAL PAY-TV SERVICE.**

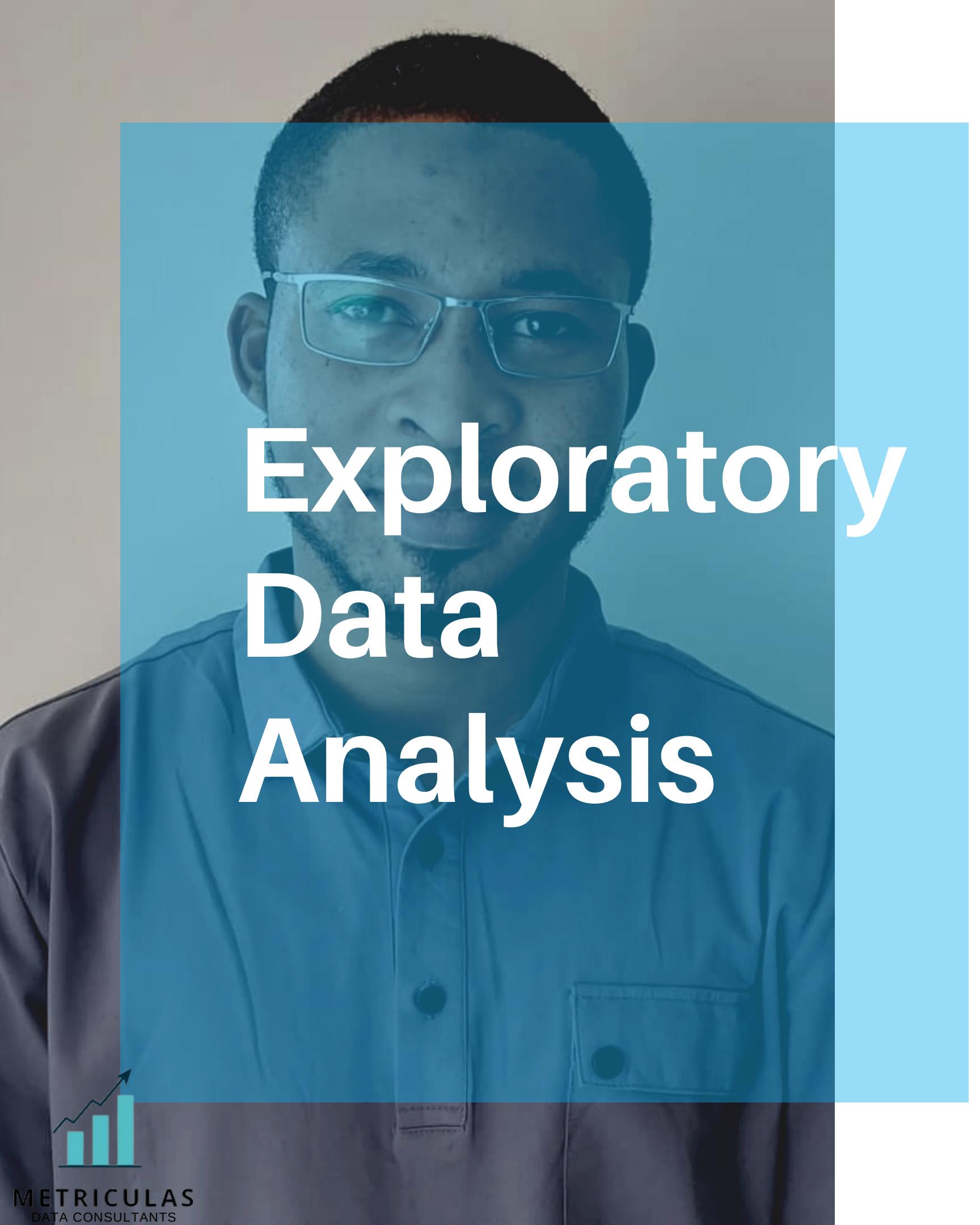
- ▶ RECOMMENDER SYSTEMS
- ▶ CONTENT-BASED FILTERING
- ▶ COLLABORATIVE FILTERING



PROBLEM STATEMENT

- To create a **content based** and **collaborative based filtering** movie recommender solution.





Exploratory Data Analysis



EXPLORATORY DATA ANALYSIS

DATA BREAKDOWN

TRAIN

10,000,038
ENTRIES

USER ID

MOVIE ID

RATING

TIMESTAMP

MOVIES

62,423
ENTRIES

MOVIE ID

TITLE

GENRE

IMDB

27,278
ENTRIES

MOVIE
ID

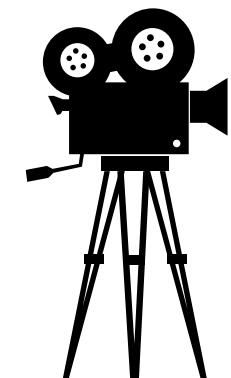
TITLE
CAST

DIRECTOR

RUNTIME

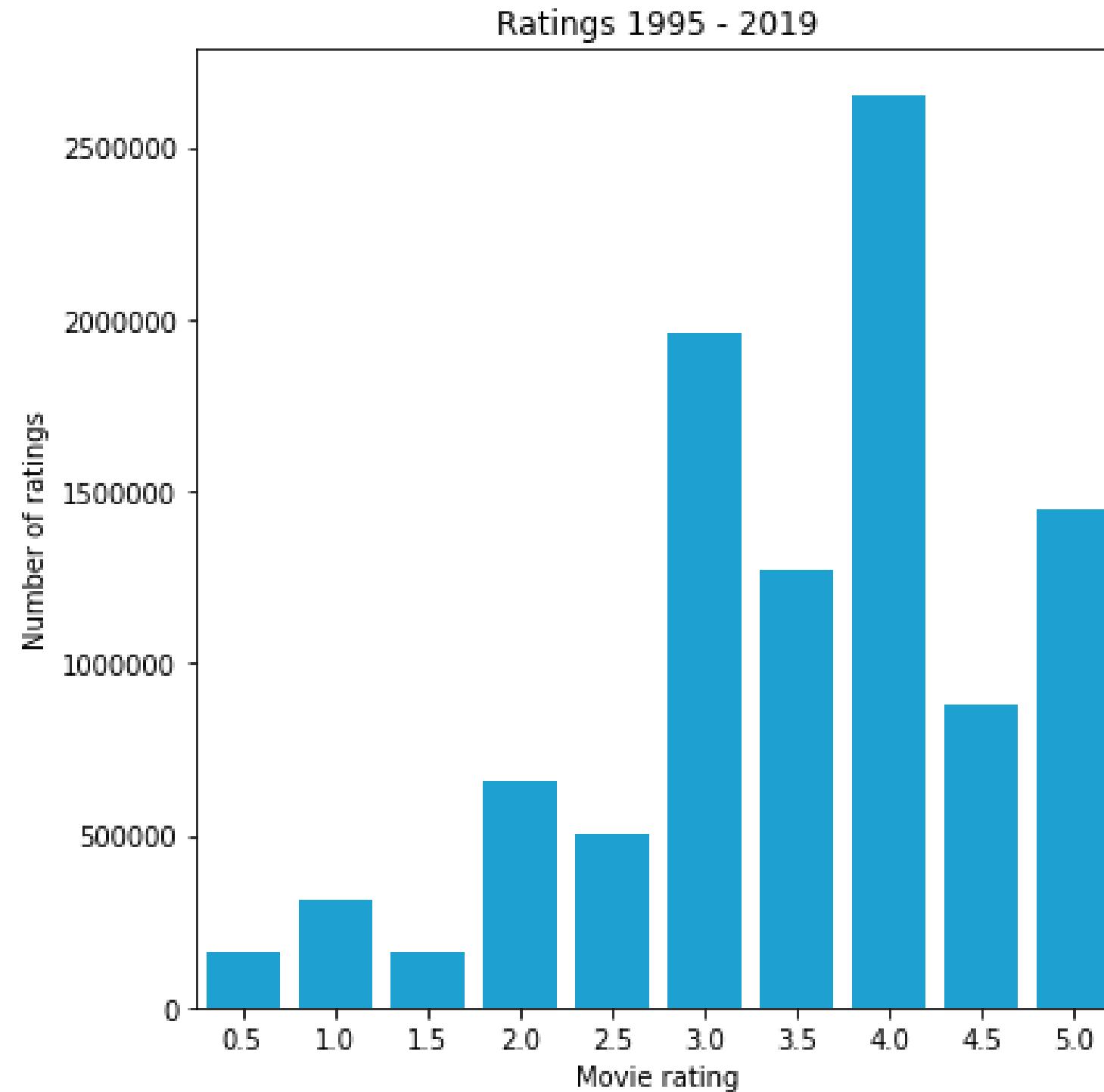
BUDGET

PLOT
KEYWORD



- Our target is recommend '[Ratings](#)' of users given their history

EXPLORATORY DATA ANALYSIS



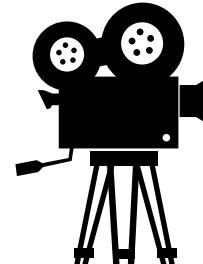
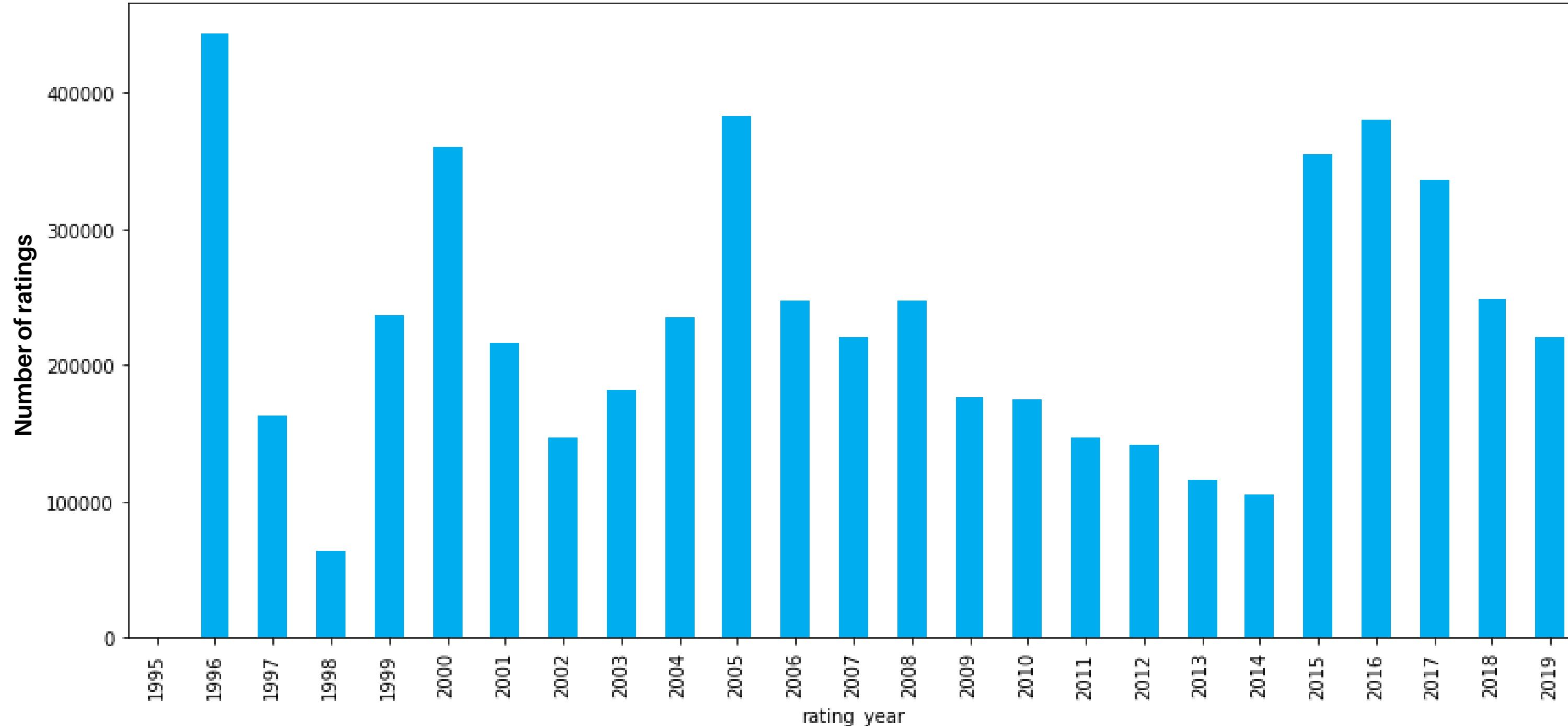
USER-RATING COUNT

- Ratings range from 0 - 5.0
- Most users rated 4.0



EXPLORATORY DATA ANALYSIS

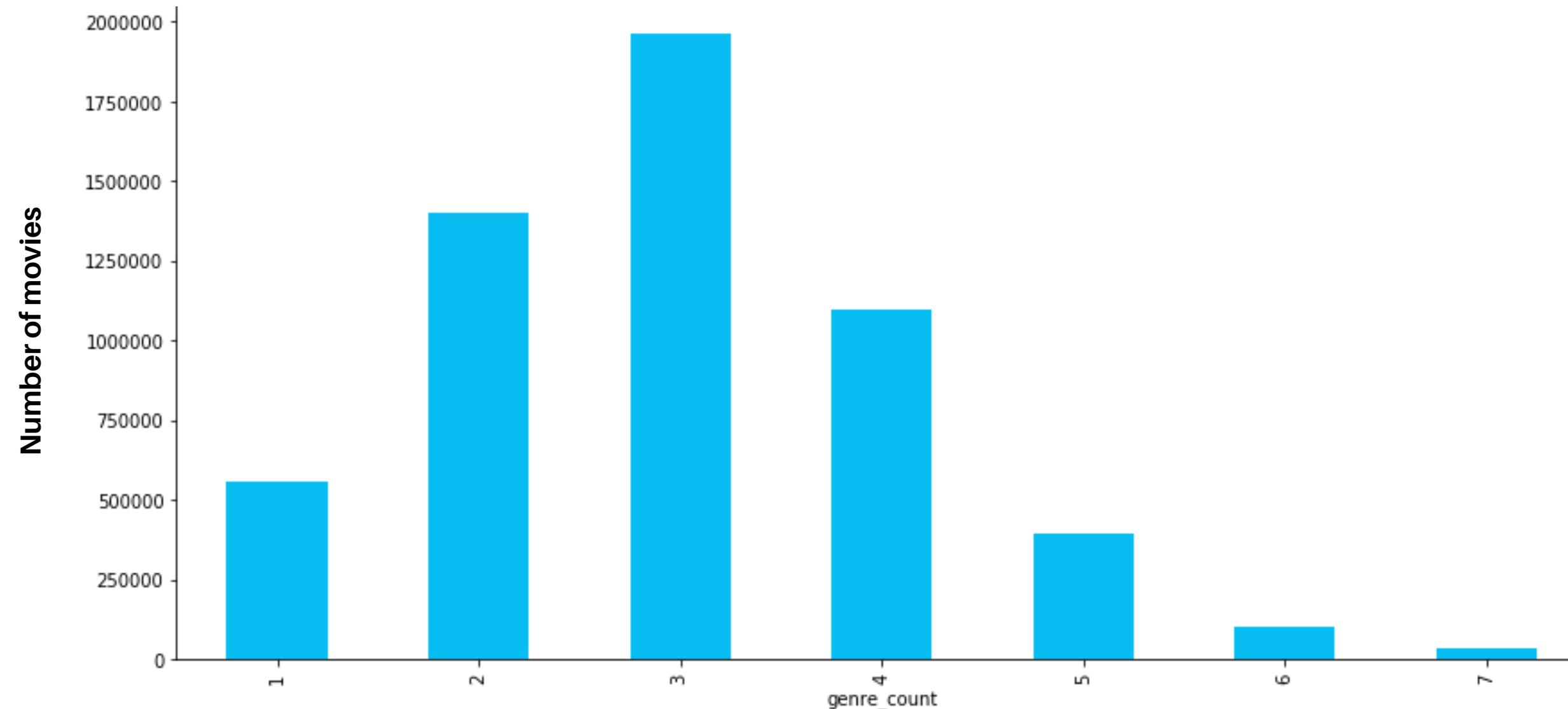
RATING YEAR (1995-2019)



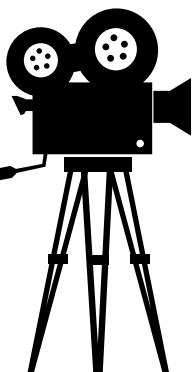
- The last 5 years accumalatively had the most ratings

EXPLORATORY DATA ANALYSIS

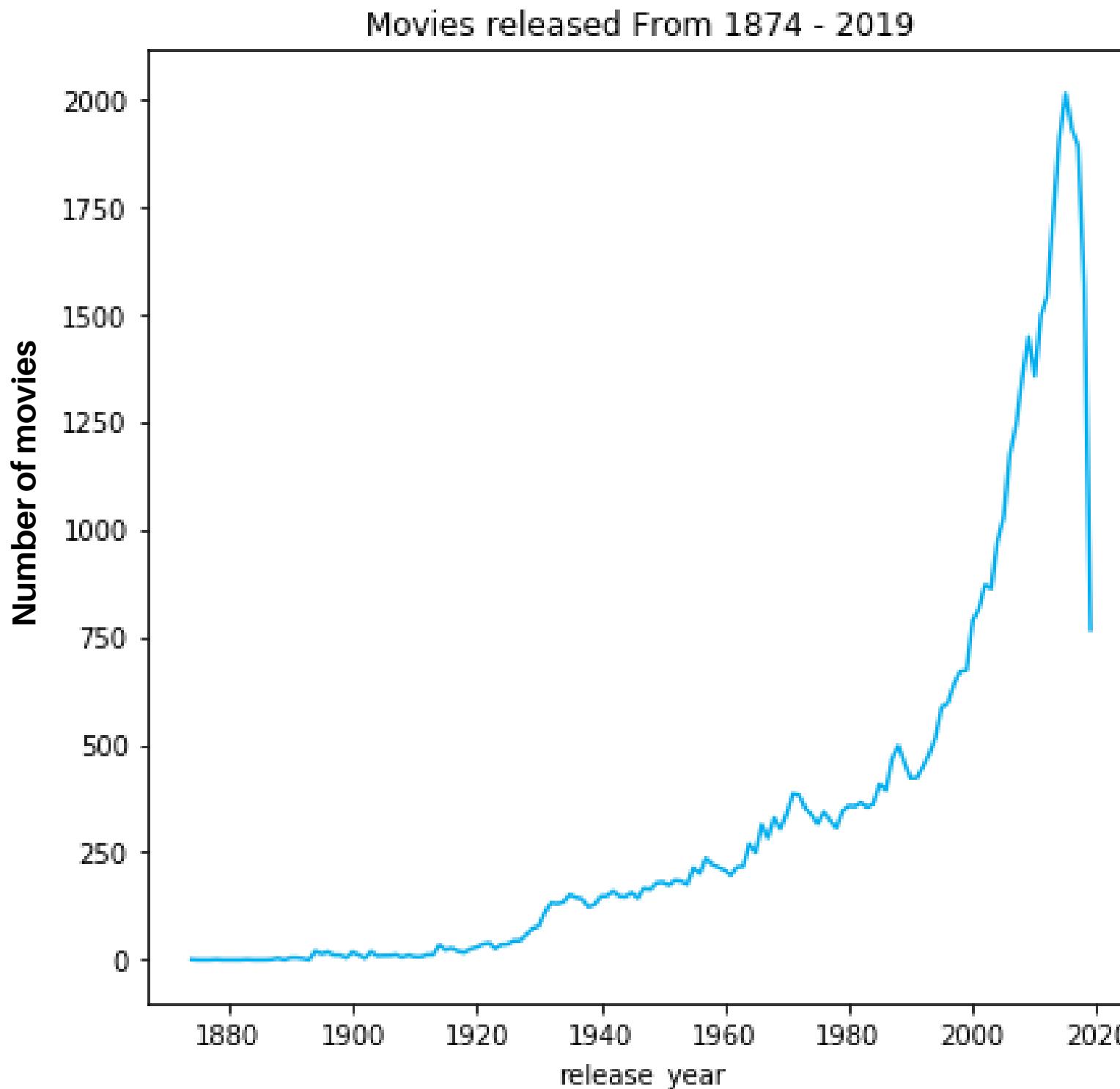
GENRE PER MOVIE



- From the image, movies with 2 - 5 genres seem to have more ratings.



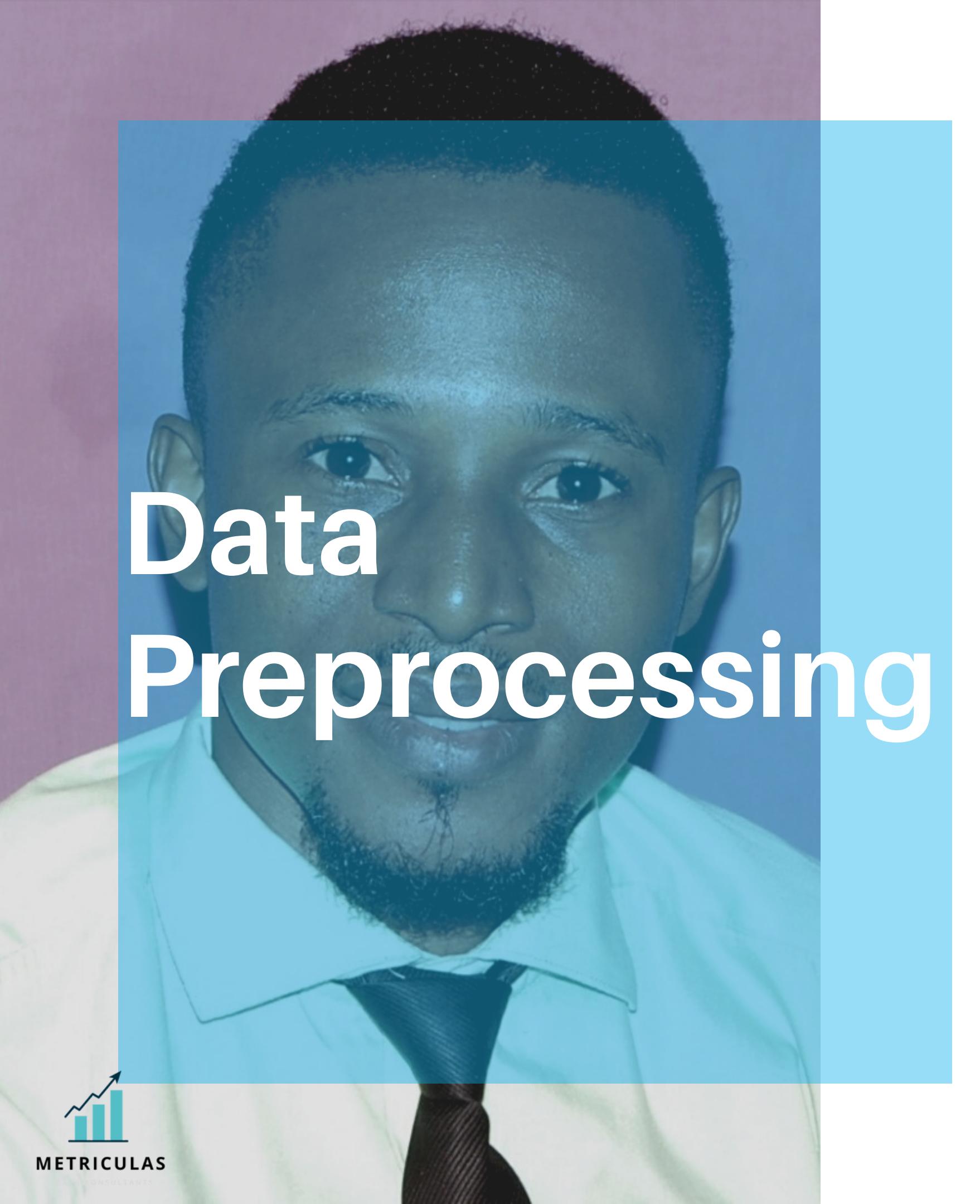
EXPLORATORY DATA ANALYSIS



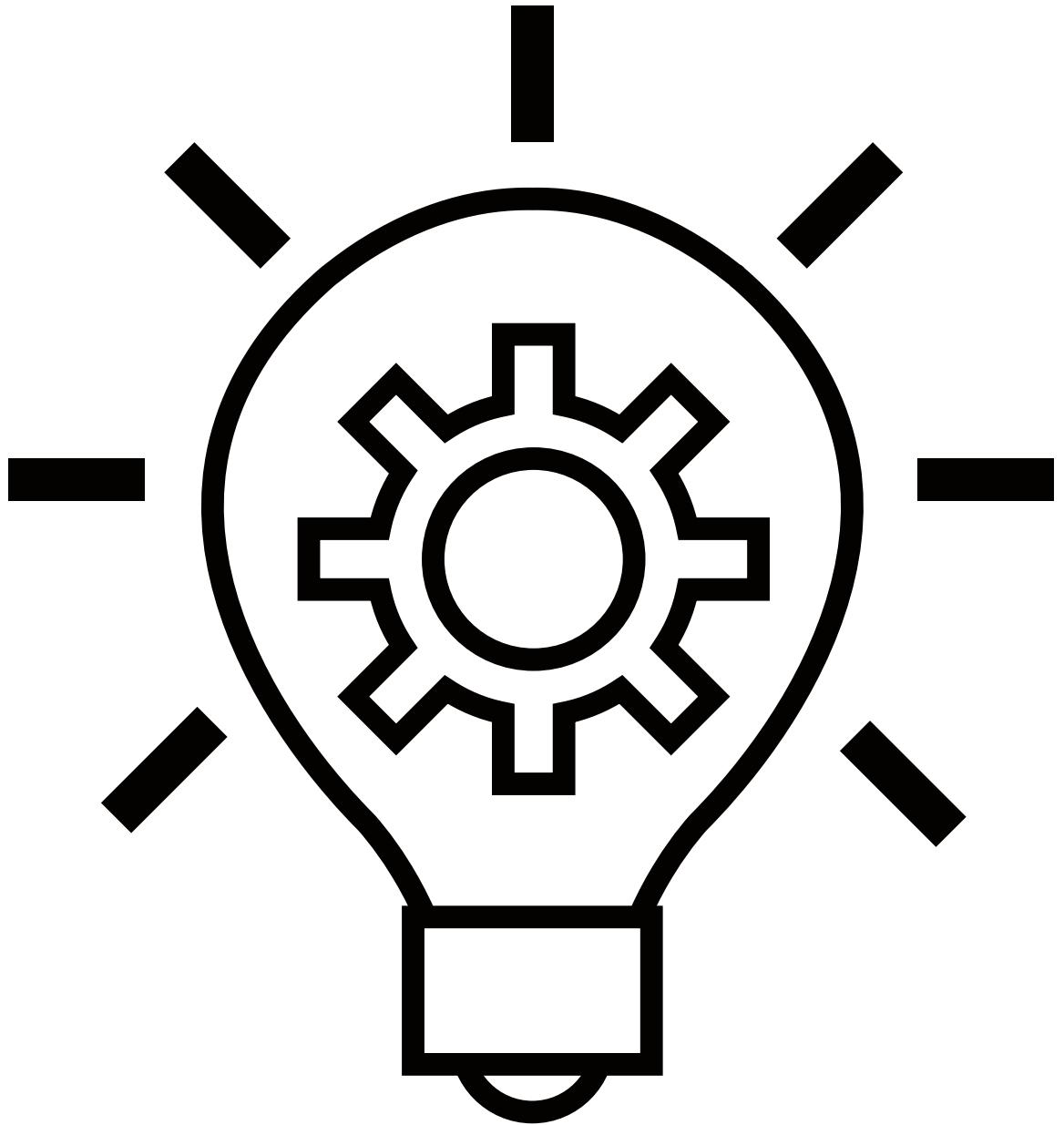
MOVIES RELEASE DATE

- Increase in slope till 2015





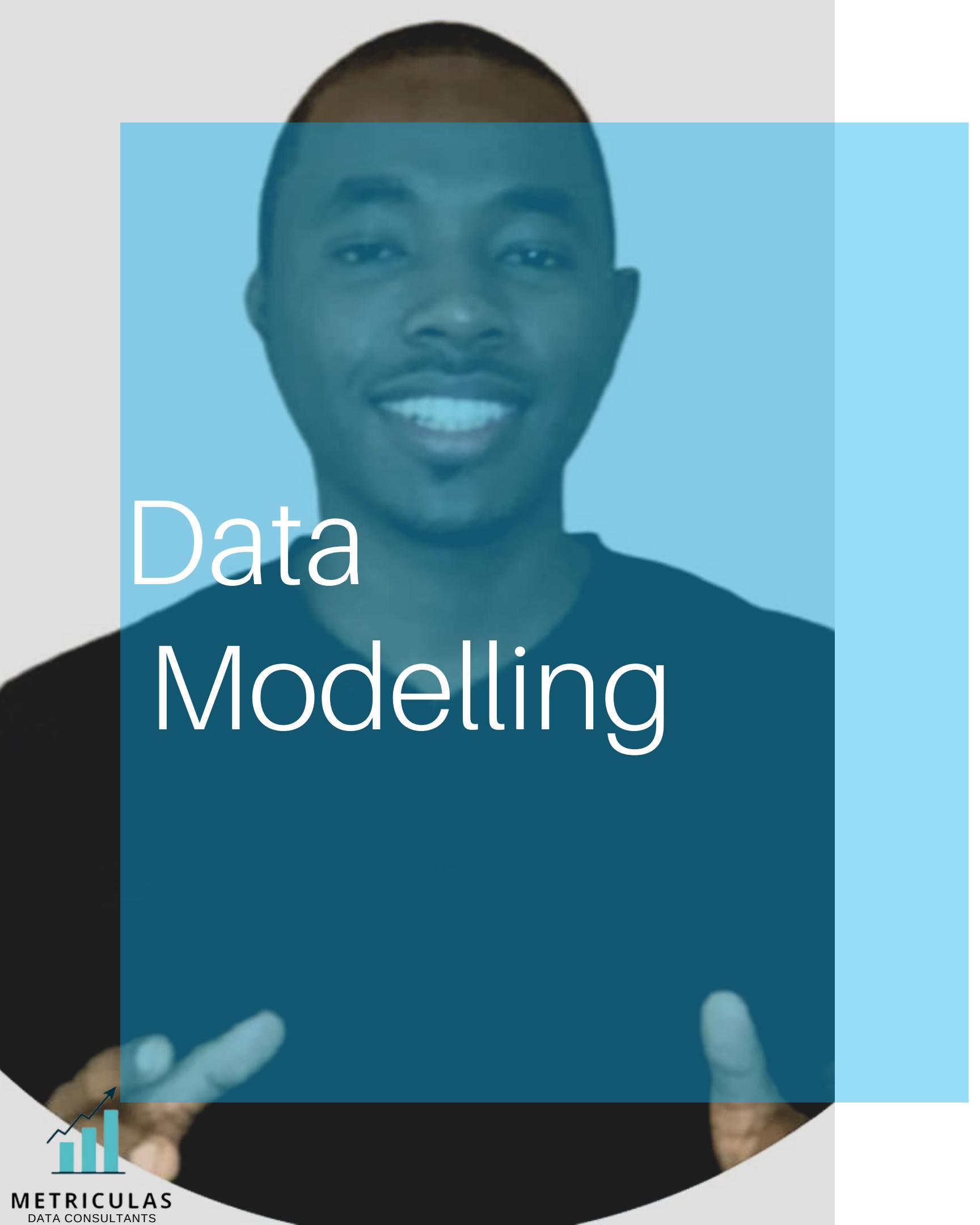
Data Preprocessing



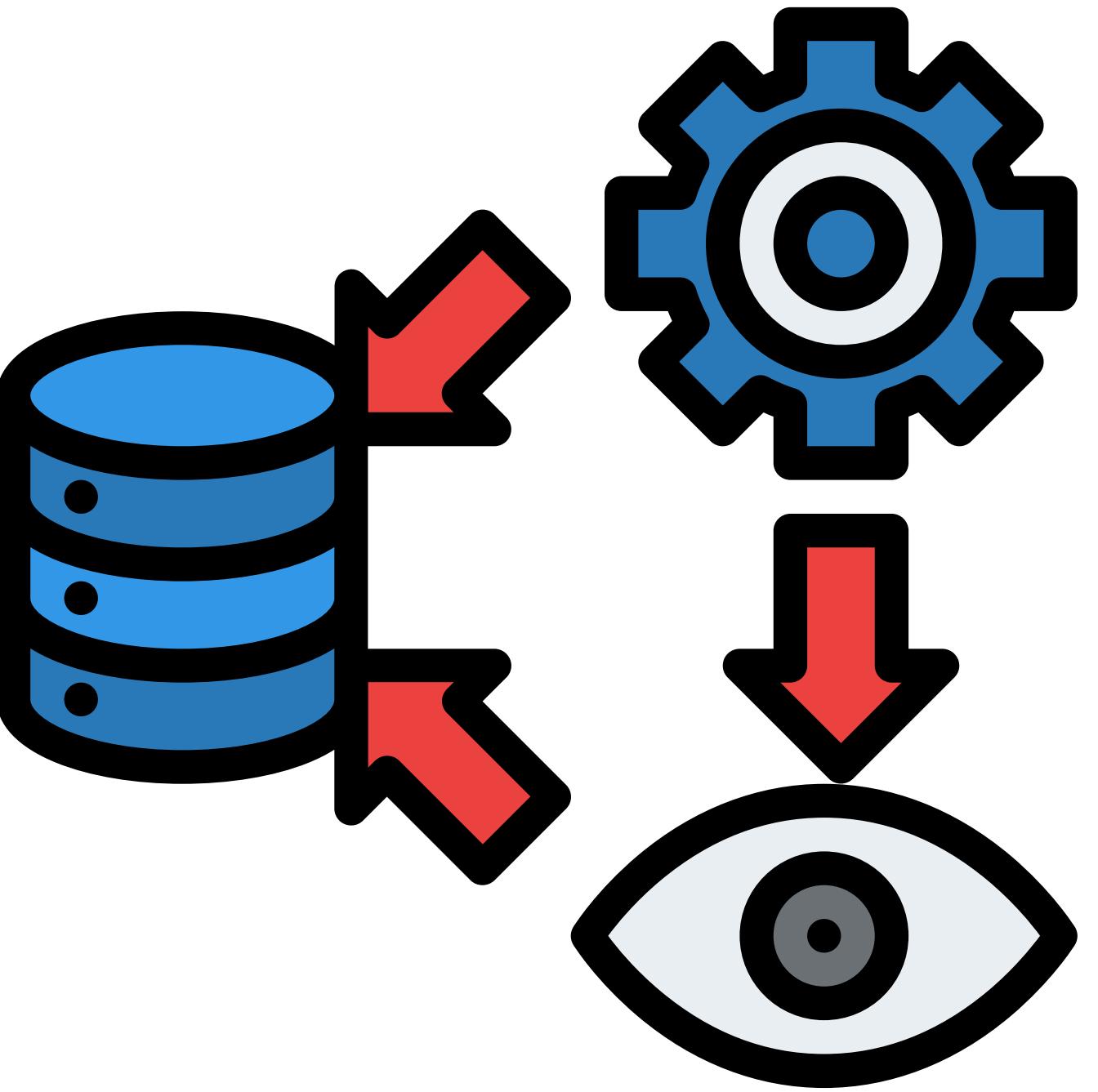
DATA PREPROCESSING



- Filtered only movies with over 2000 ratings
- Created new feature of genres list and count
- Dropped the timestamp column

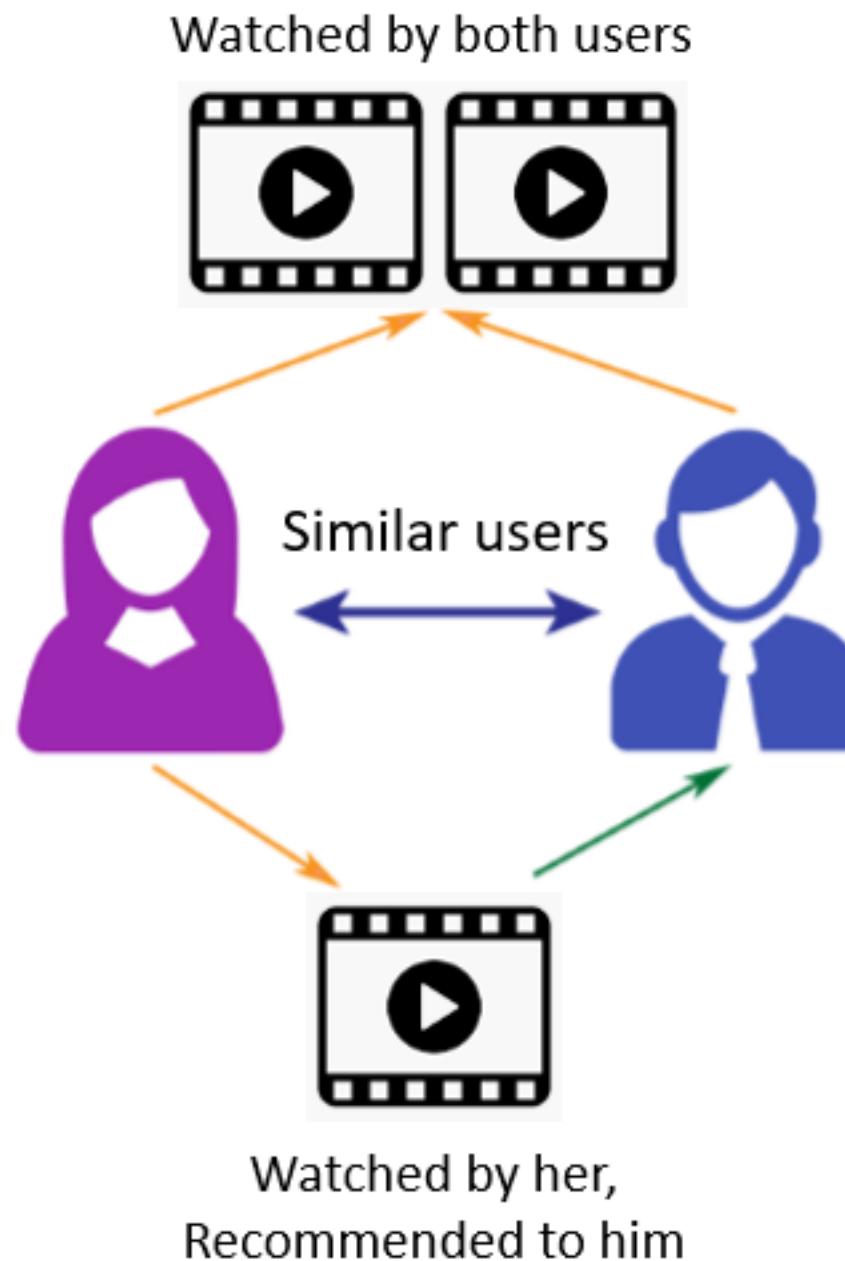


Data Modelling

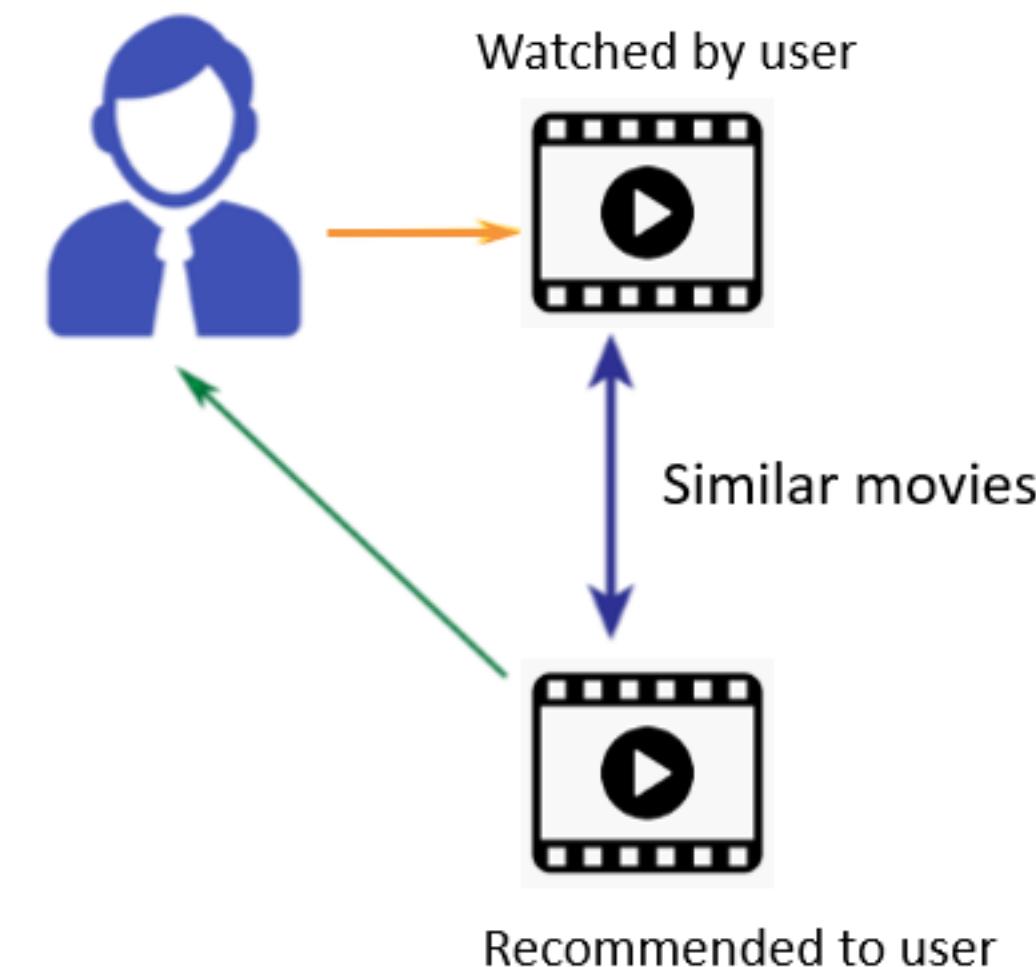


MODELS

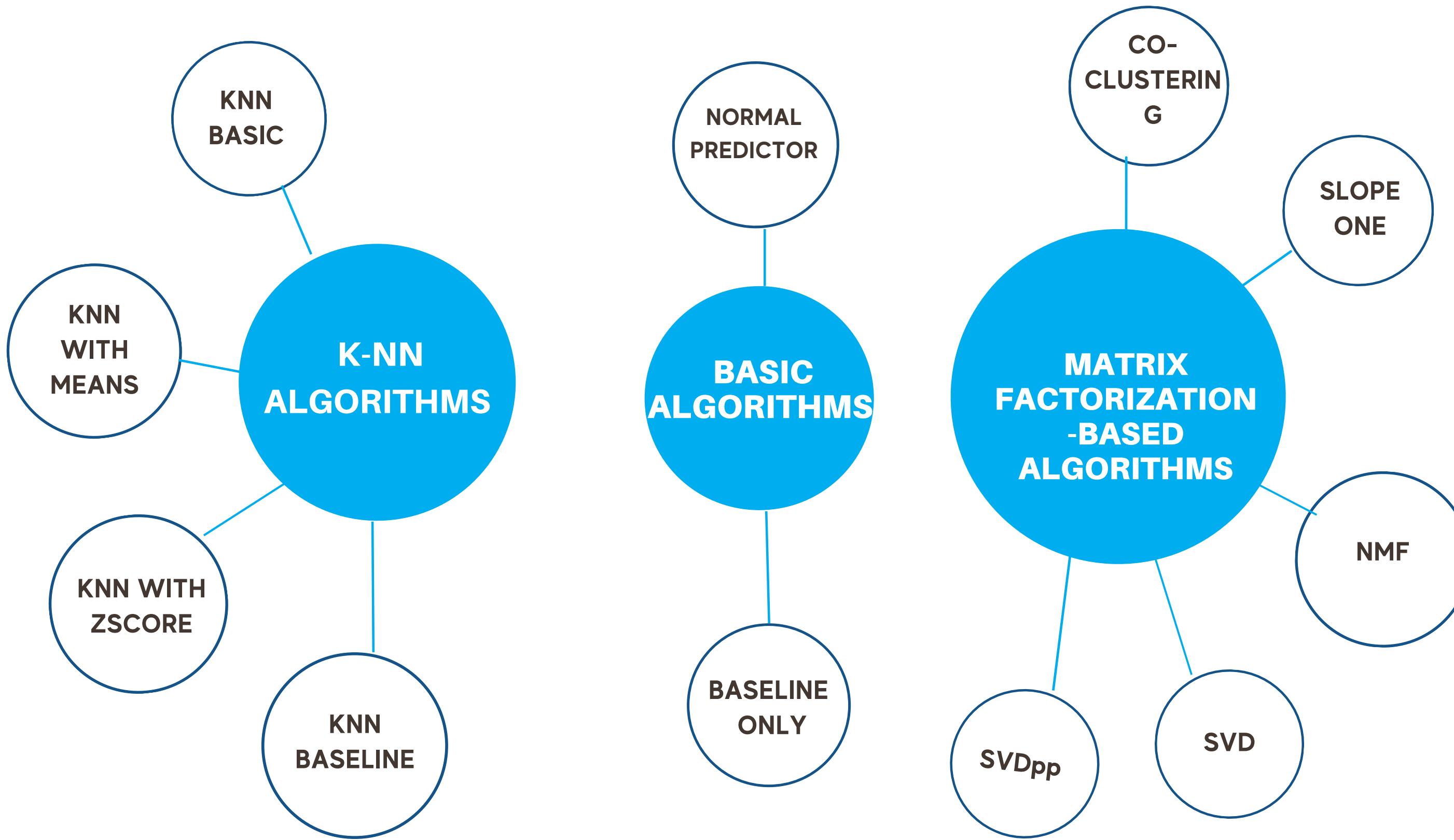
Collaborative Filtering



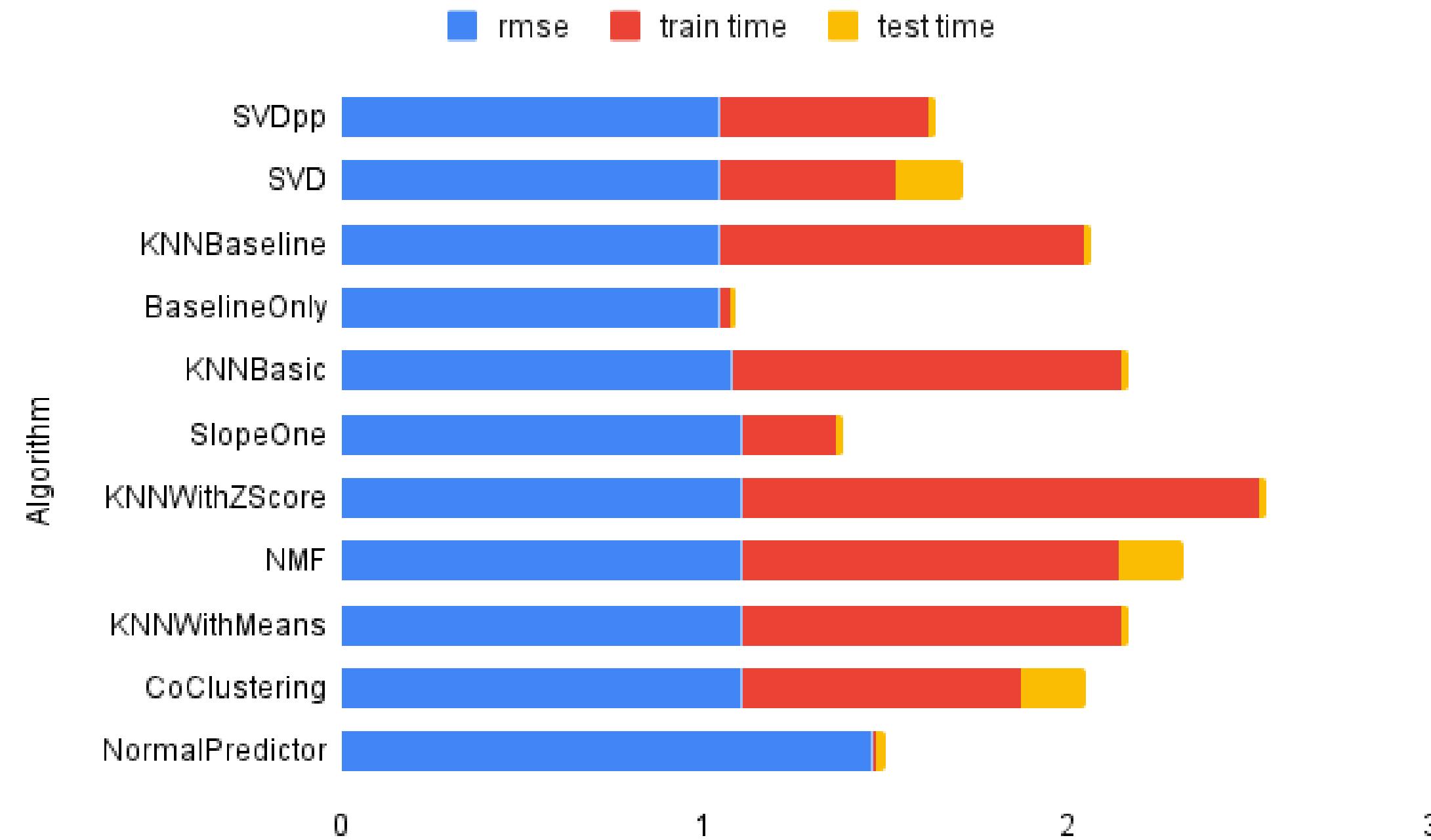
Content-Based Filtering



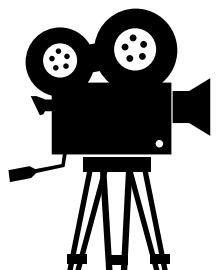
MODELS USED (COLLABORATIVE FILTERING)



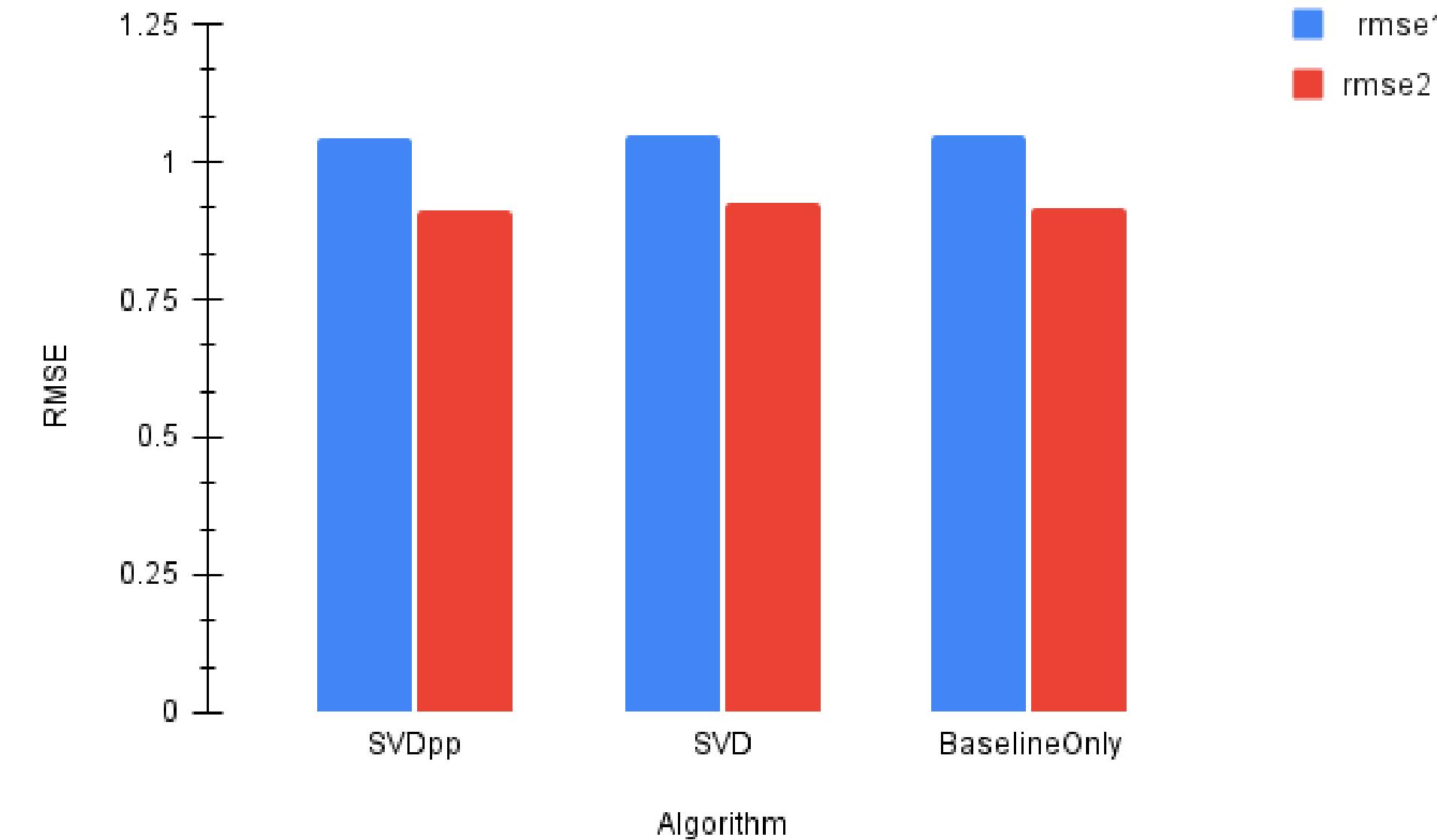
MODELS PERFORMANCE



We chose the SVD, SVDPp, BASELINE only model

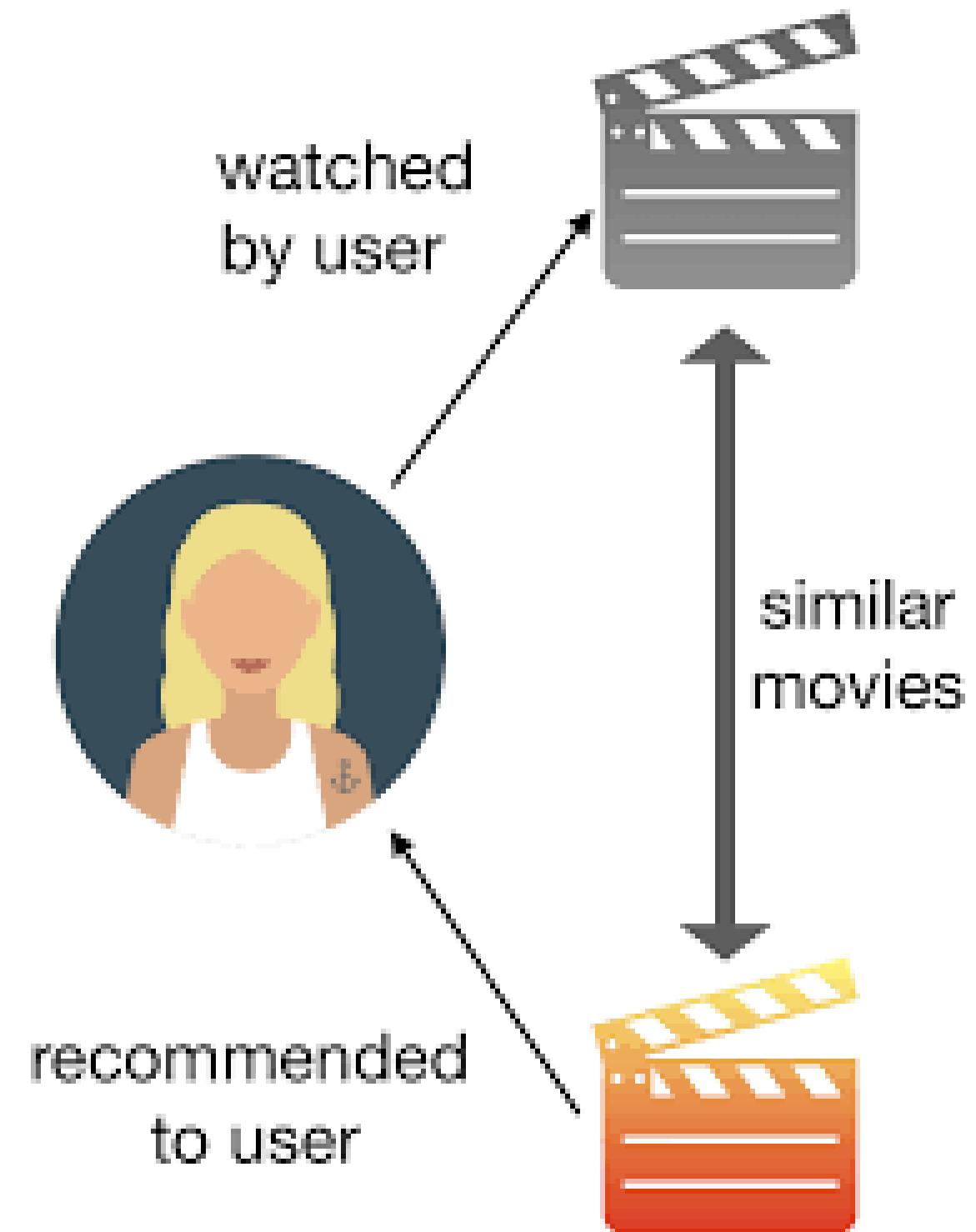


MODEL IMPROVEMENT



- Models improved with the aid of hyperparameter tuning
- SVD and performed the best with a rmse of 0.91 and kaggle score of 0.78

MODEL USED



CONTENT-BASED FILTERING

- Matrix of Pivot Table of key features
- Compute **Cosine Similarity** using **KNN**

MODEL DEPLOYMENT

Not secure | 18.203.111.196:5000 www.BANDICAM.com

Choose Option

Recommender System

Movie Recommender Engine

EXPLORE Data Science Academy Unsupervised Predict

Avengers: Endgame 2019
Da 5 Bloods 2020
Parasite 2019
No Time to Die 2020
Bad Boys for Life 2020
The King's Man 2020
Wonder Woman 1984 2020

Select an algorithm

Content Based Filtering
 Collaborative Based Filtering

Enter Your Three Favorite Movies

First Option

RECOMMENDATIONS AND CONCLUSION

RECOMMENDATION

- We need more computational power
- SVD best results with low computation time
- We recommend hybrid approach to filtering

Thank you!

Any Questions?

