

DSTV Movie Recommender System

by
STATA Consulting



Melania Makombe
CEO



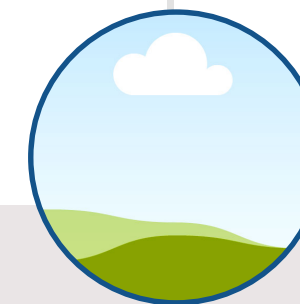
Abubakar Sadiq
Data Explorer



Michael Kanu
API, Github
Specialist



Nuraddeen Isah
Juniour Data
Engineer

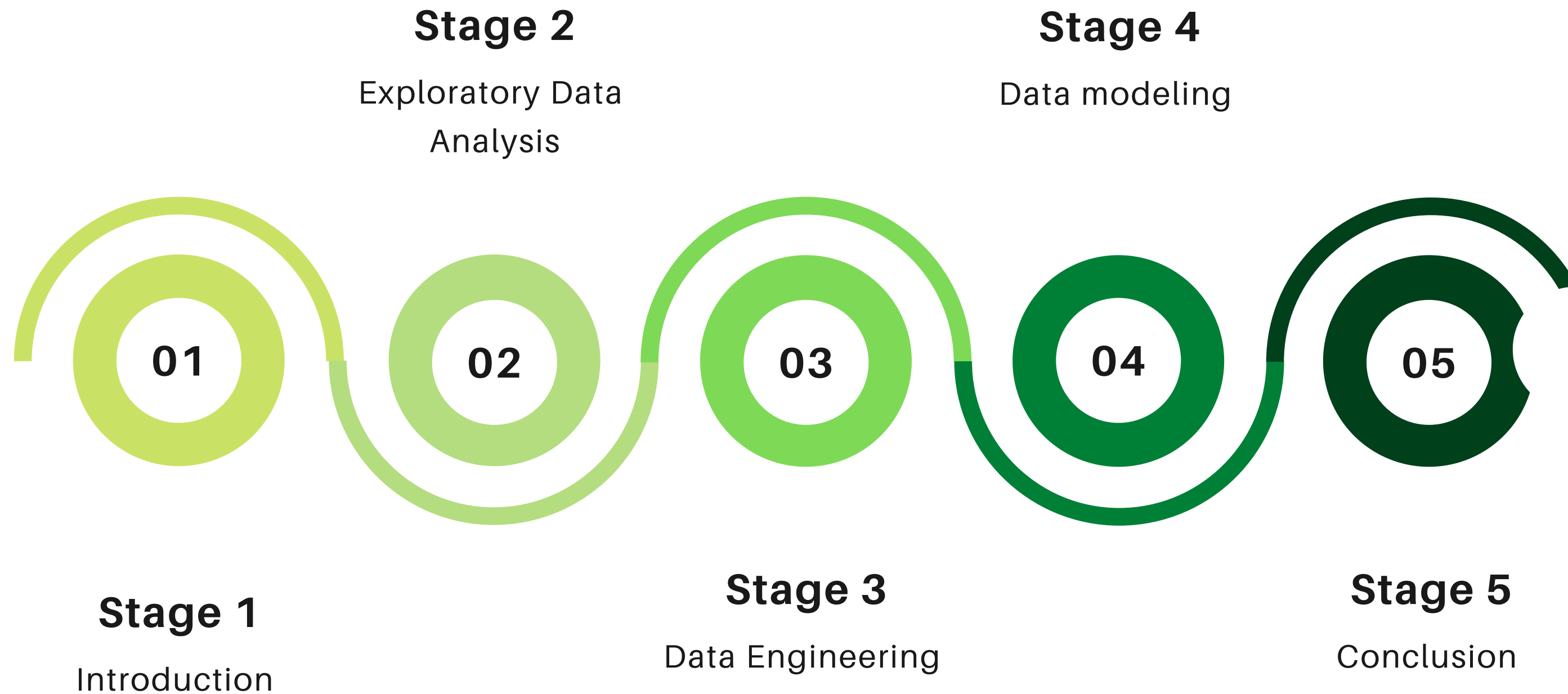


Christian DivineFavour
Senior Modeling Engineer

ABOUT US

- Operating since 1995
- Big clients include Tesla and Netflix
- Website: www.strata.co.za

TABLE OF CONTENTS

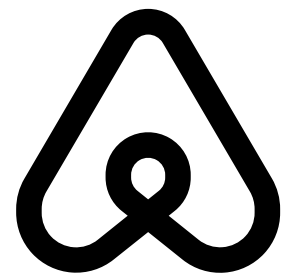


INTRODUCTION

- 69% consumers prefer video streaming over traditional pay-TV service.
- Create a collaborative and content based recomender sytstem

amazon

NETFLIX



MOVIE RECOMMENDER SYSTEM ROADMAP

Data collection

Download data from Kaggle

Exploratory data science

Understanding the data mophology

Data Preprocessing

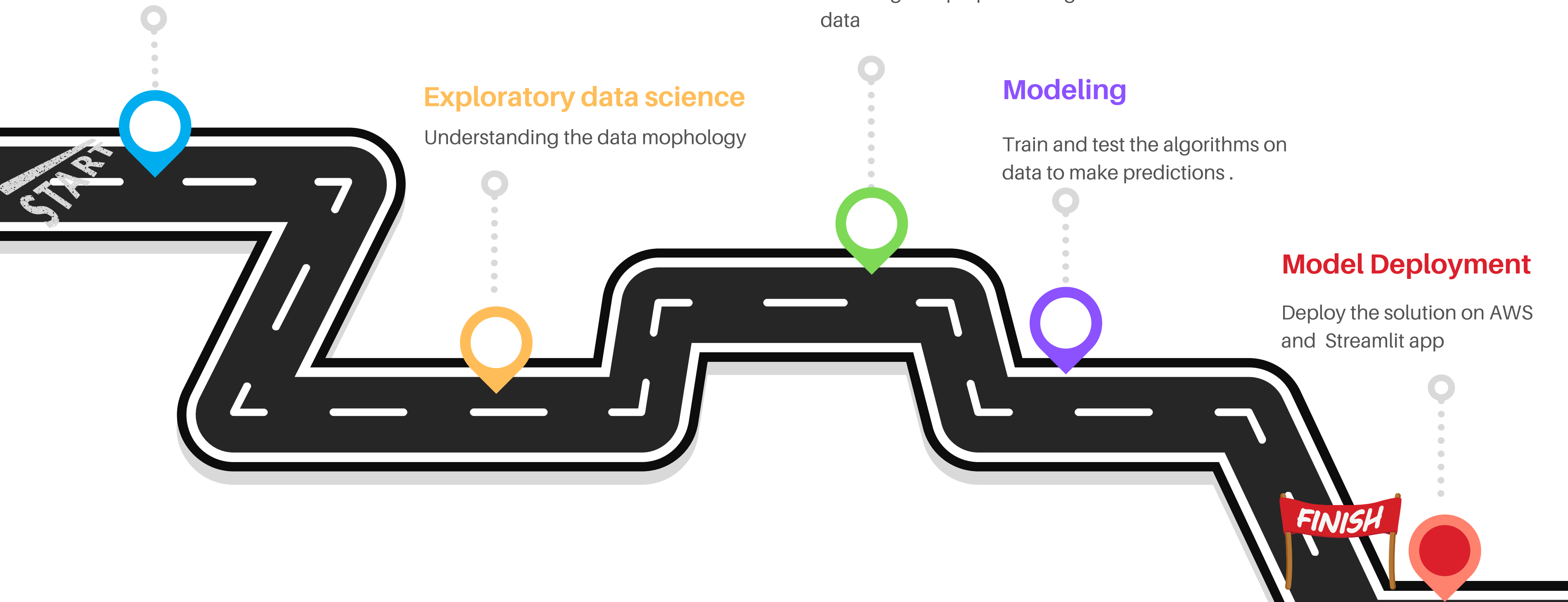
Cleaning and preprocessing the data

Modeling

Train and test the algorithms on data to make predictions .

Model Deployment

Deploy the solution on AWS and Streamlit app



Exploratory Data Analysis

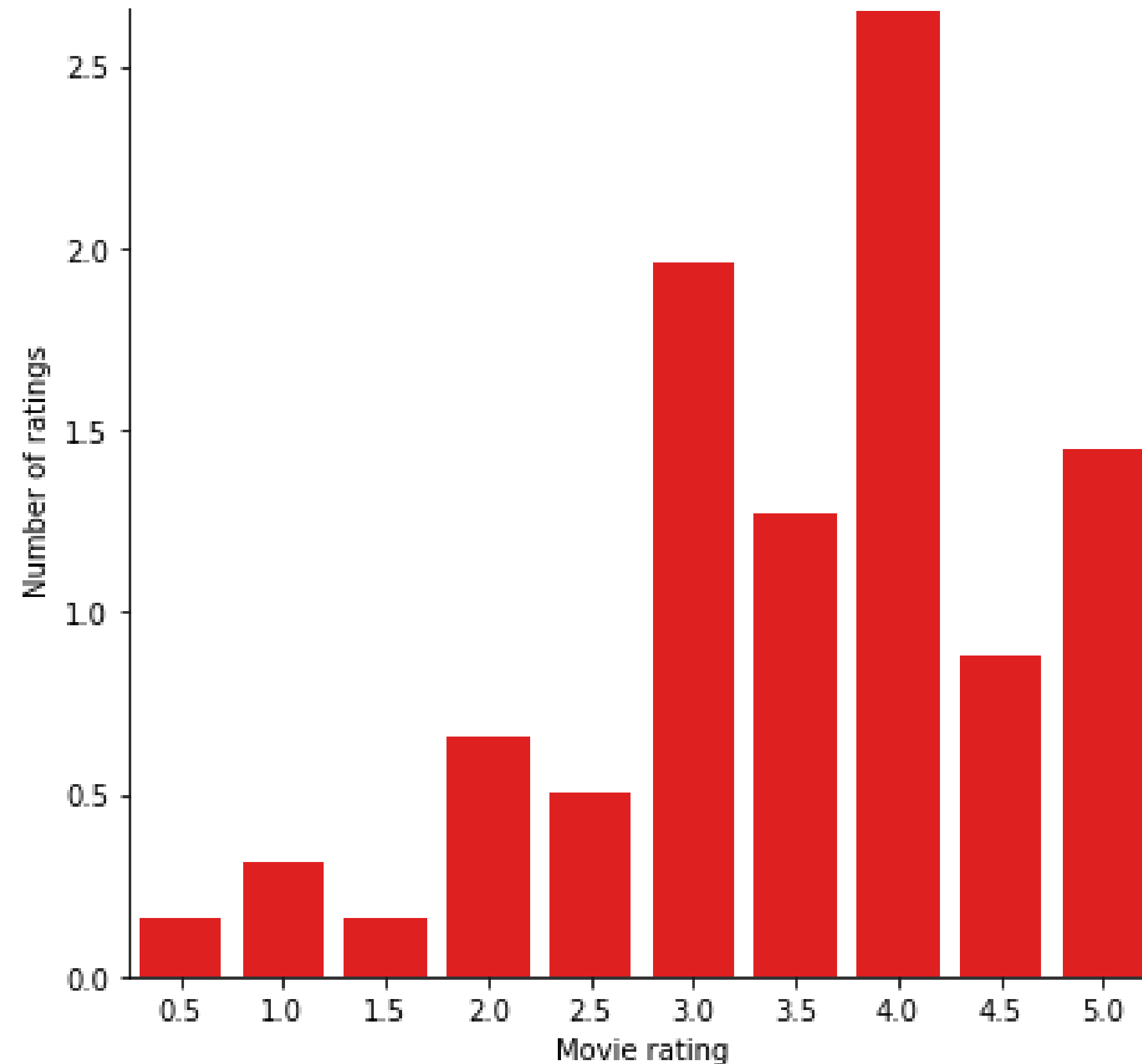


EXPLORATORY DATA ANALYSIS

DATA BREAKDOWN

- Test Data - 10000038 rows and 4 features
- Movies Data - 62423 rows and 3 features
- Imbd Data - 27278 rows and 6 features
- Our target is recommend 'Ratings' of users given their history

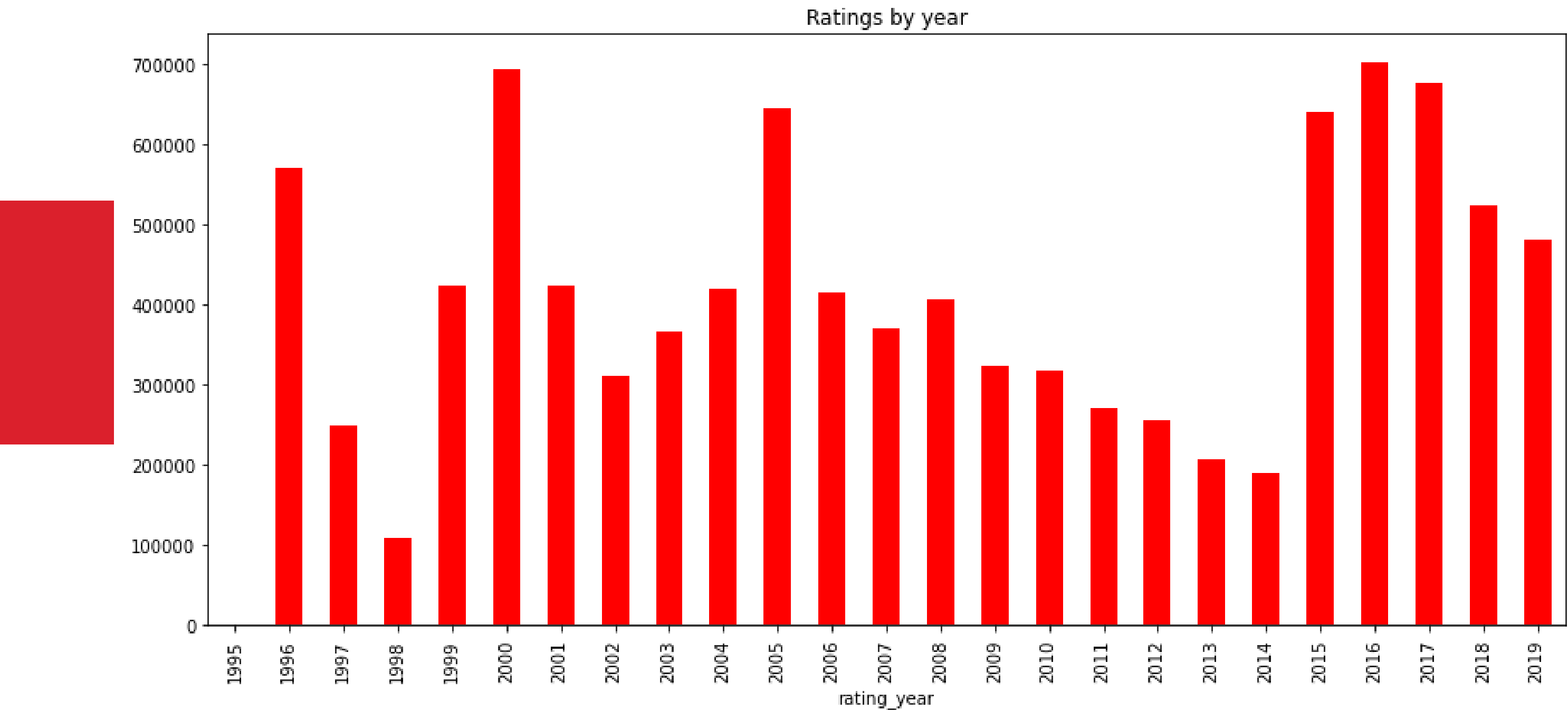
EXPLORATORY DATA ANALYSIS



USER-RATING COUNT

- Ratings range from 0 - 5
- Most users rated 4

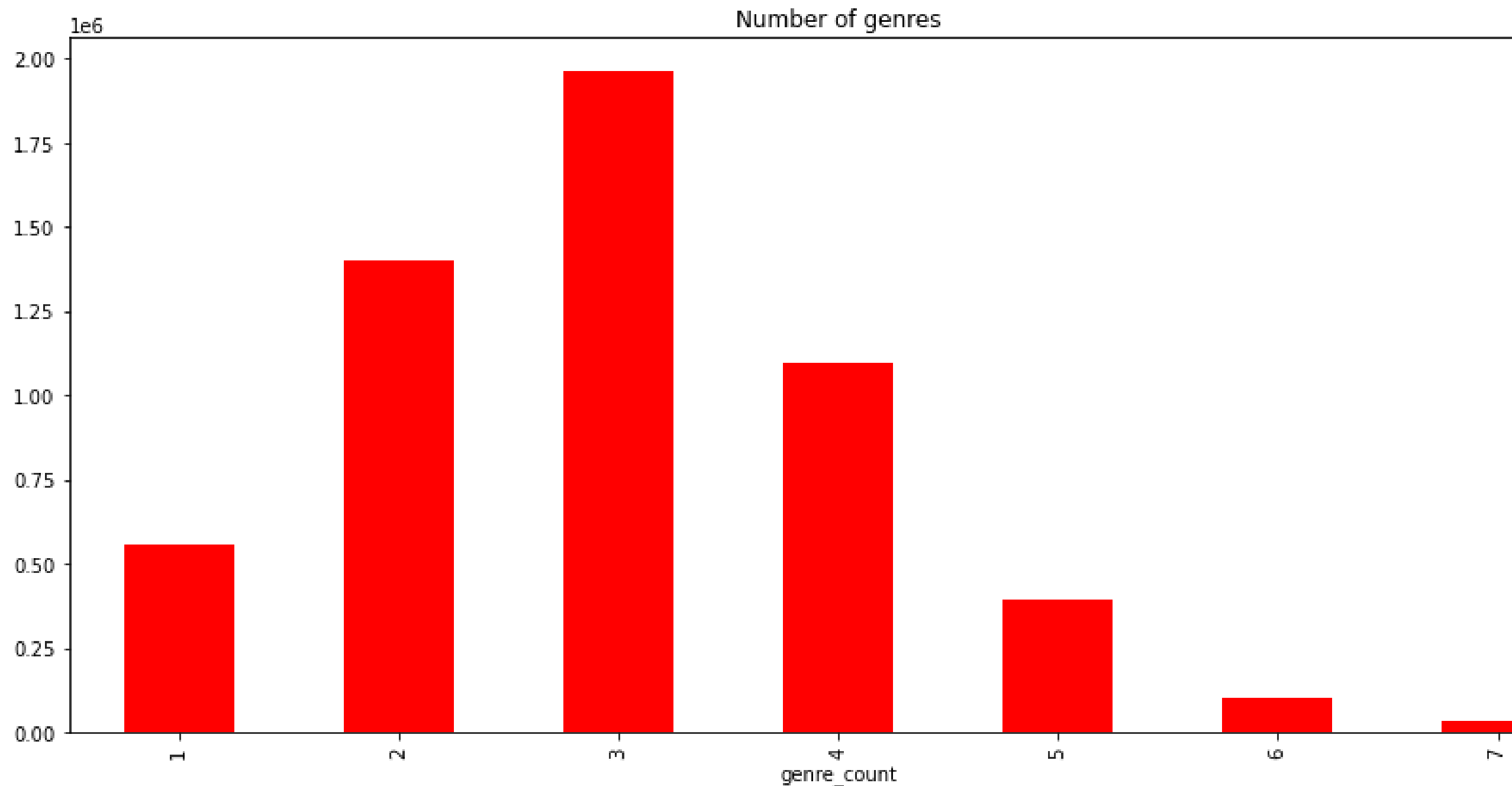
EXPLORATORY DATA ANALYSIS



RATING YEAR

- Between 1995 - 2019
- The last 5 years accumulatively having 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

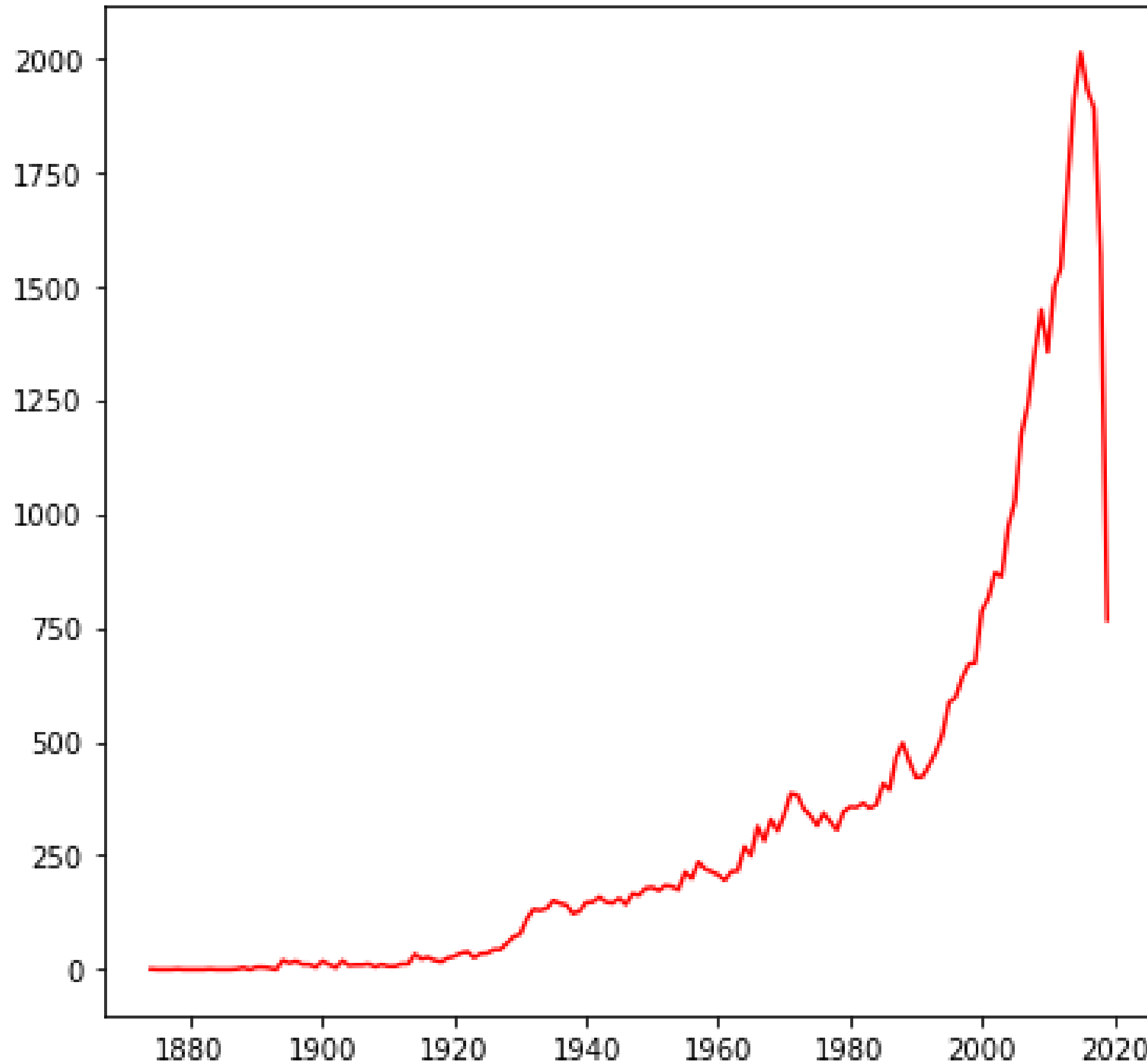


POPULAR GENRES

Words such as Sci Fi,
Action, Adventure are
prevalent

EXPLORATORY DATA ANALYSIS

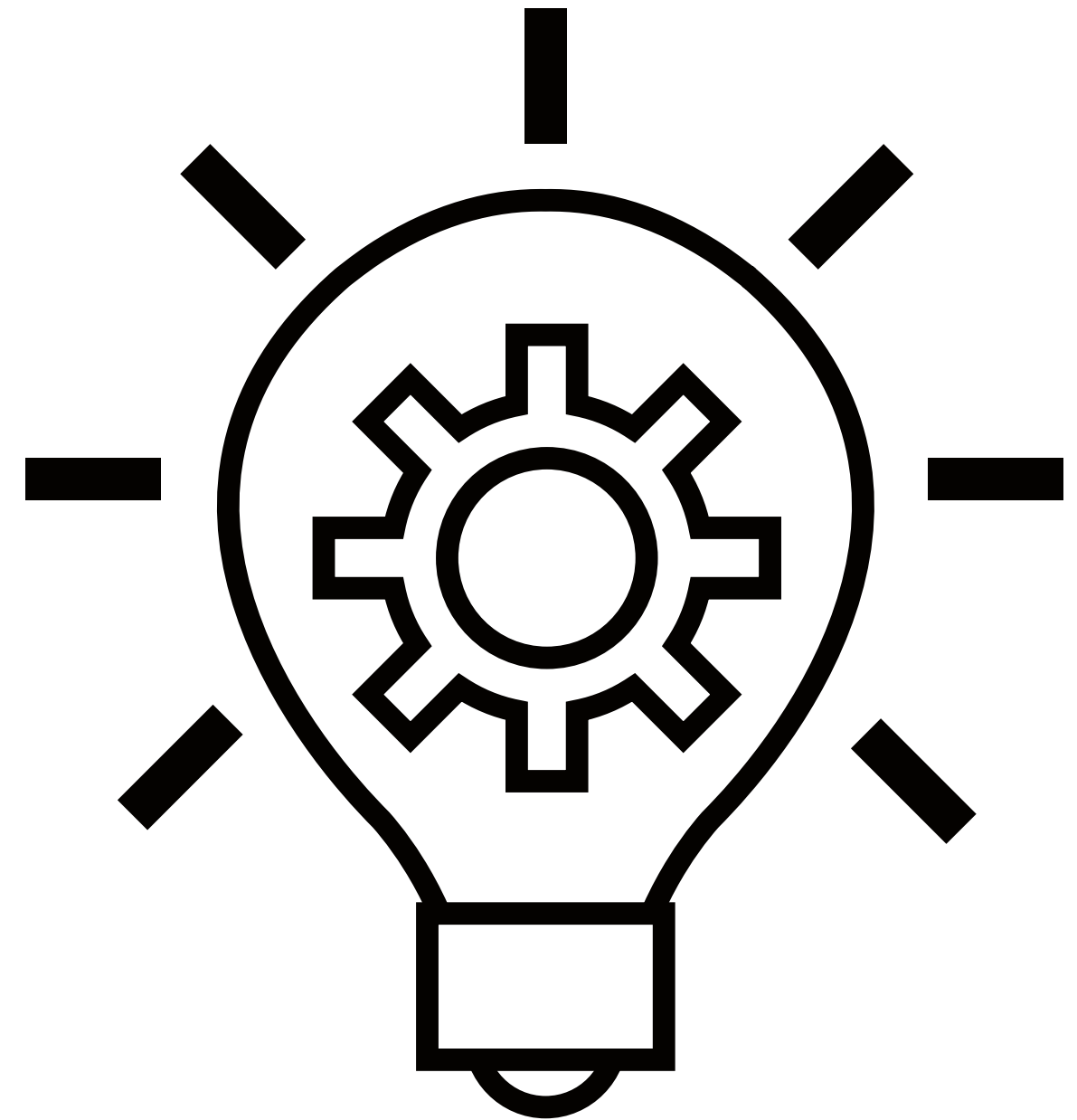
Movies released From 1874 - 2019



MOVIES
RELEASE DATE

- Increase in slope till 2015

Data Preprocessing



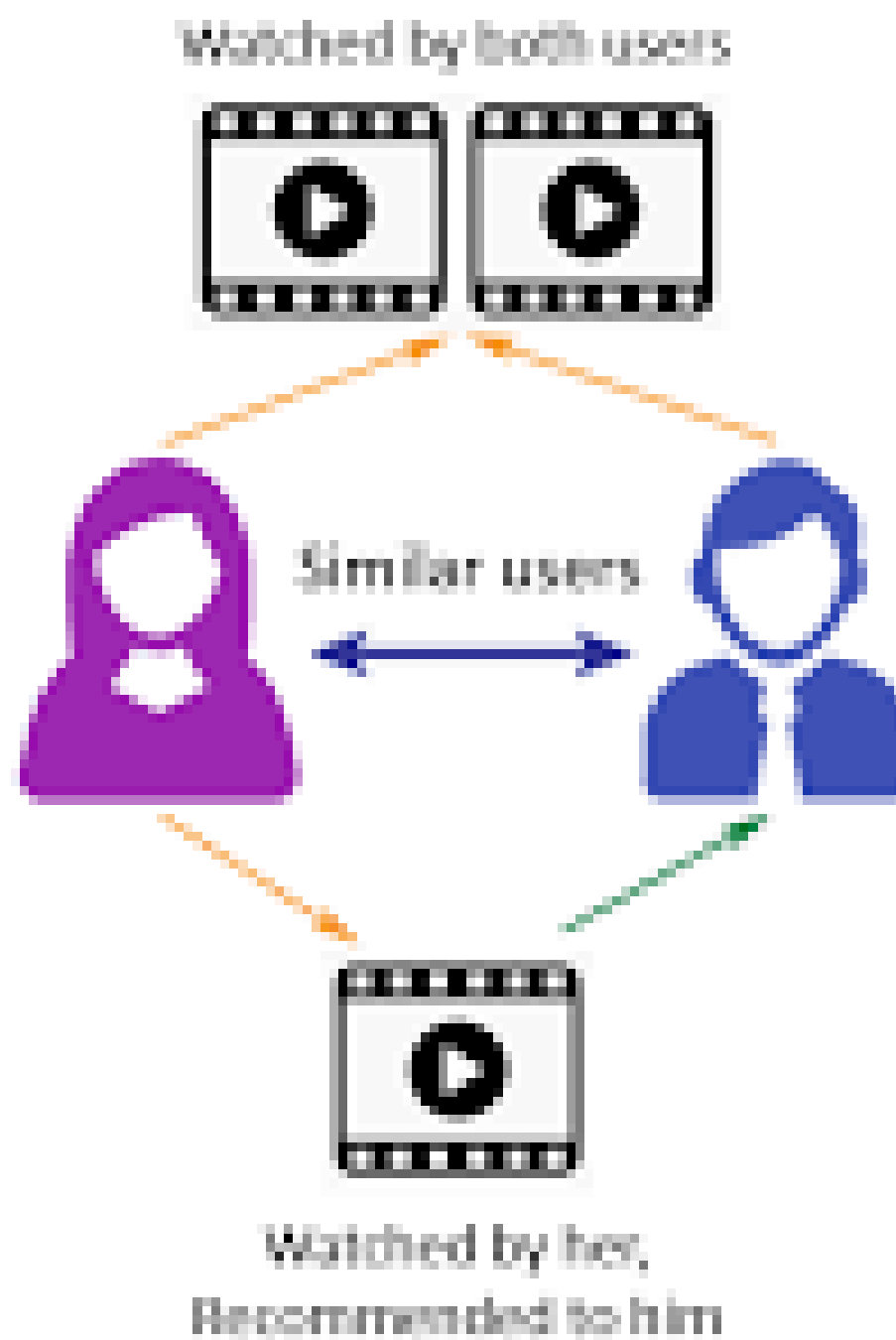
DATA PREPROCESSING



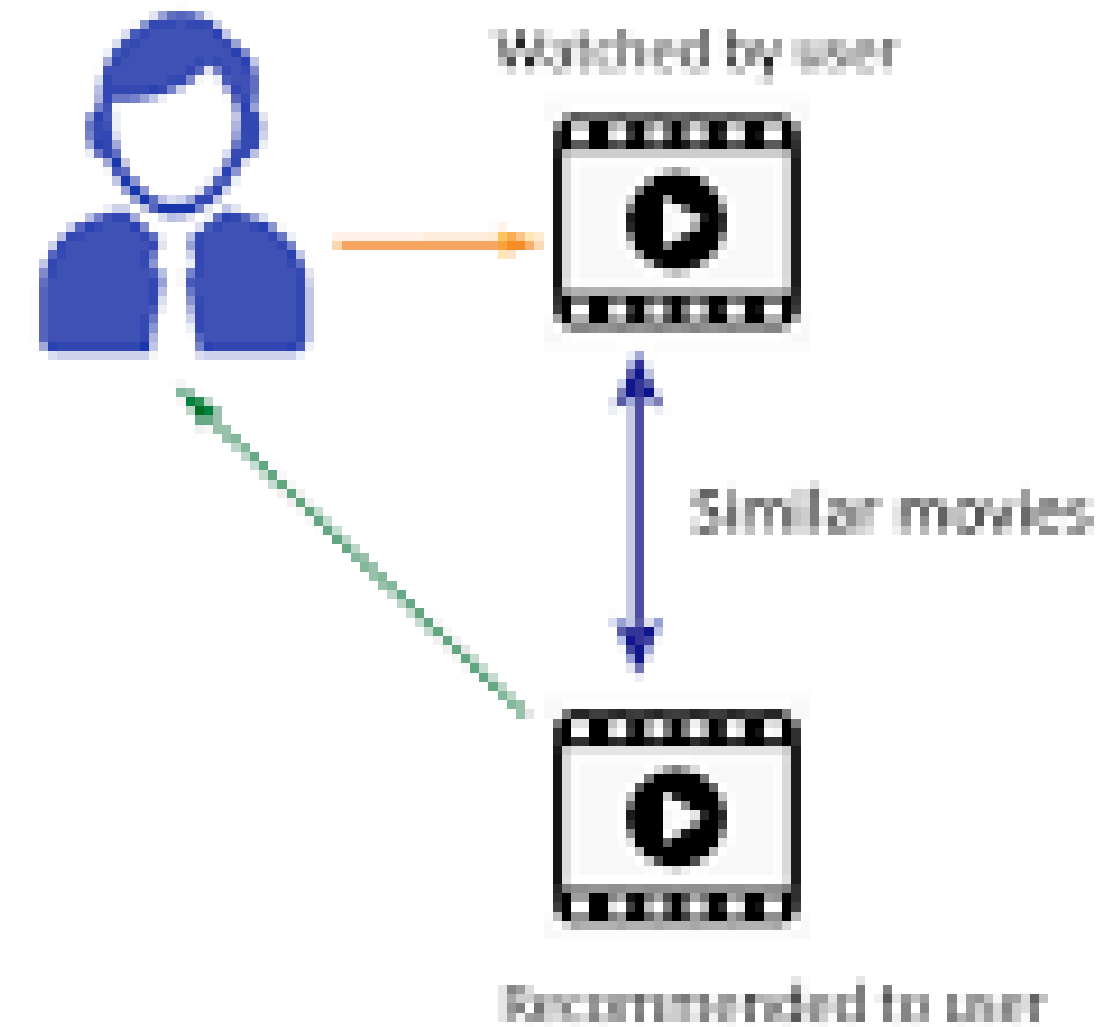
- Filtered only movies with over 2000 ratings
- Created new feature of genres list and count

Modeling

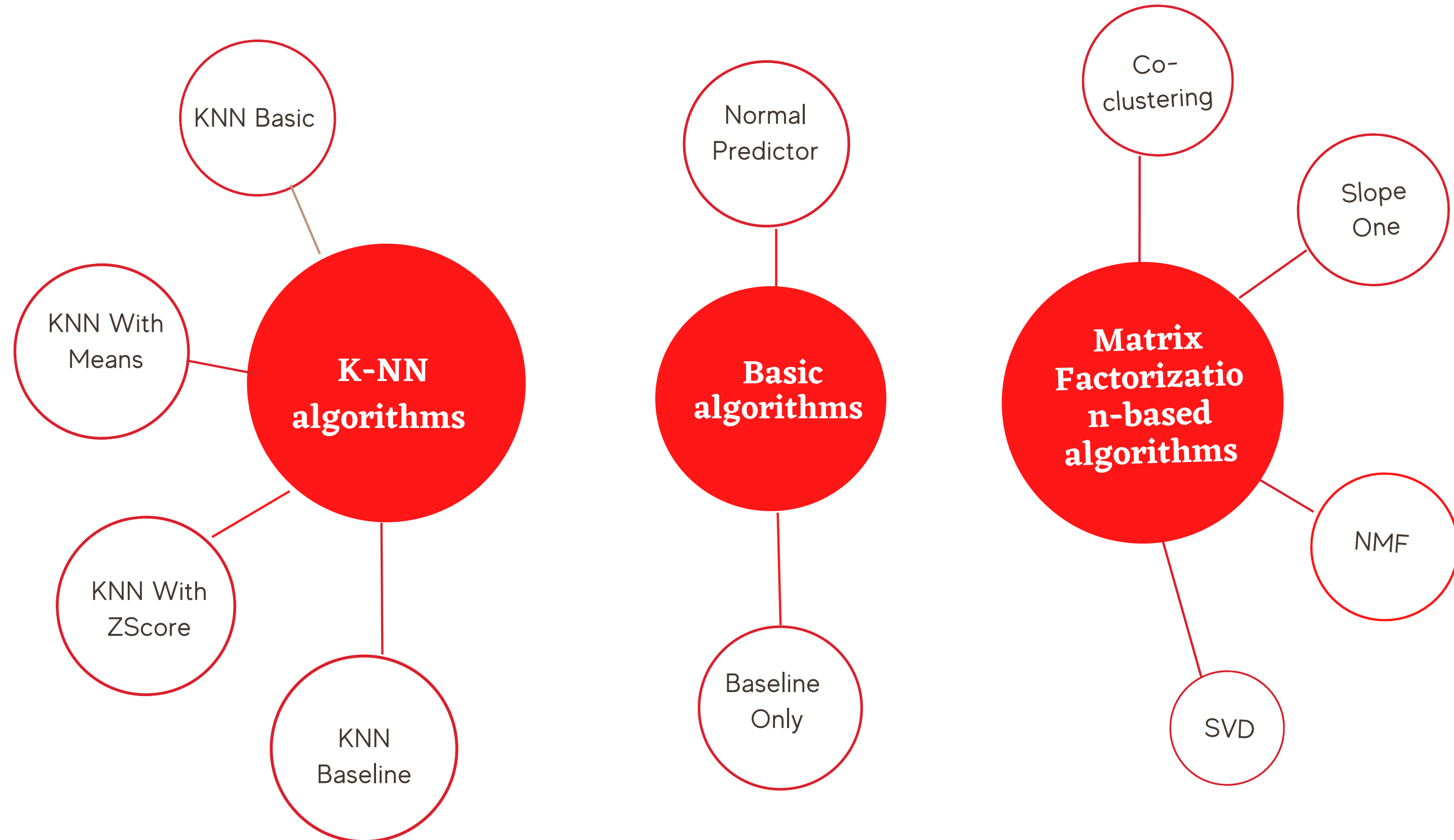
Collaborative Filtering



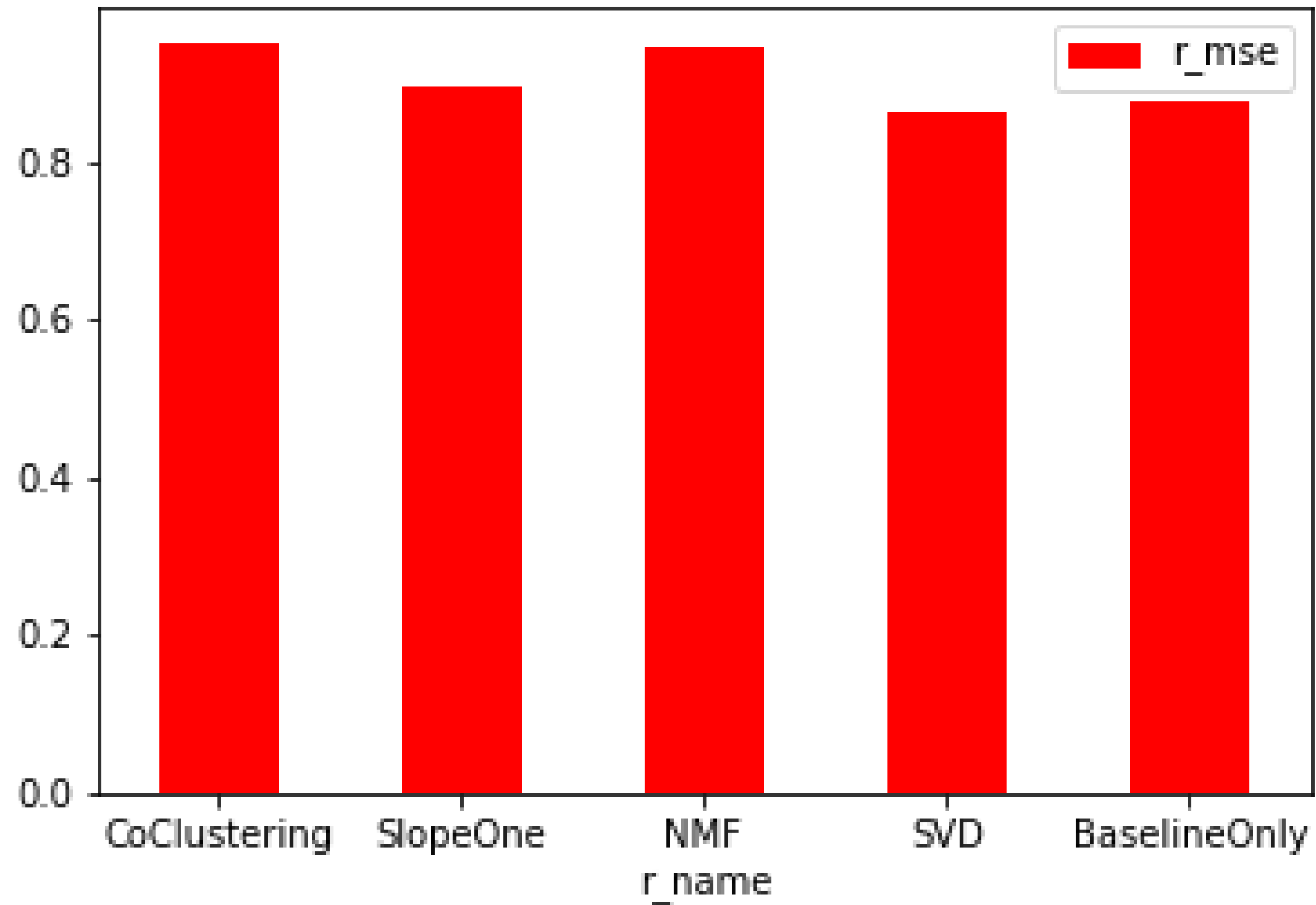
Content-Based Filtering



MODELS USED



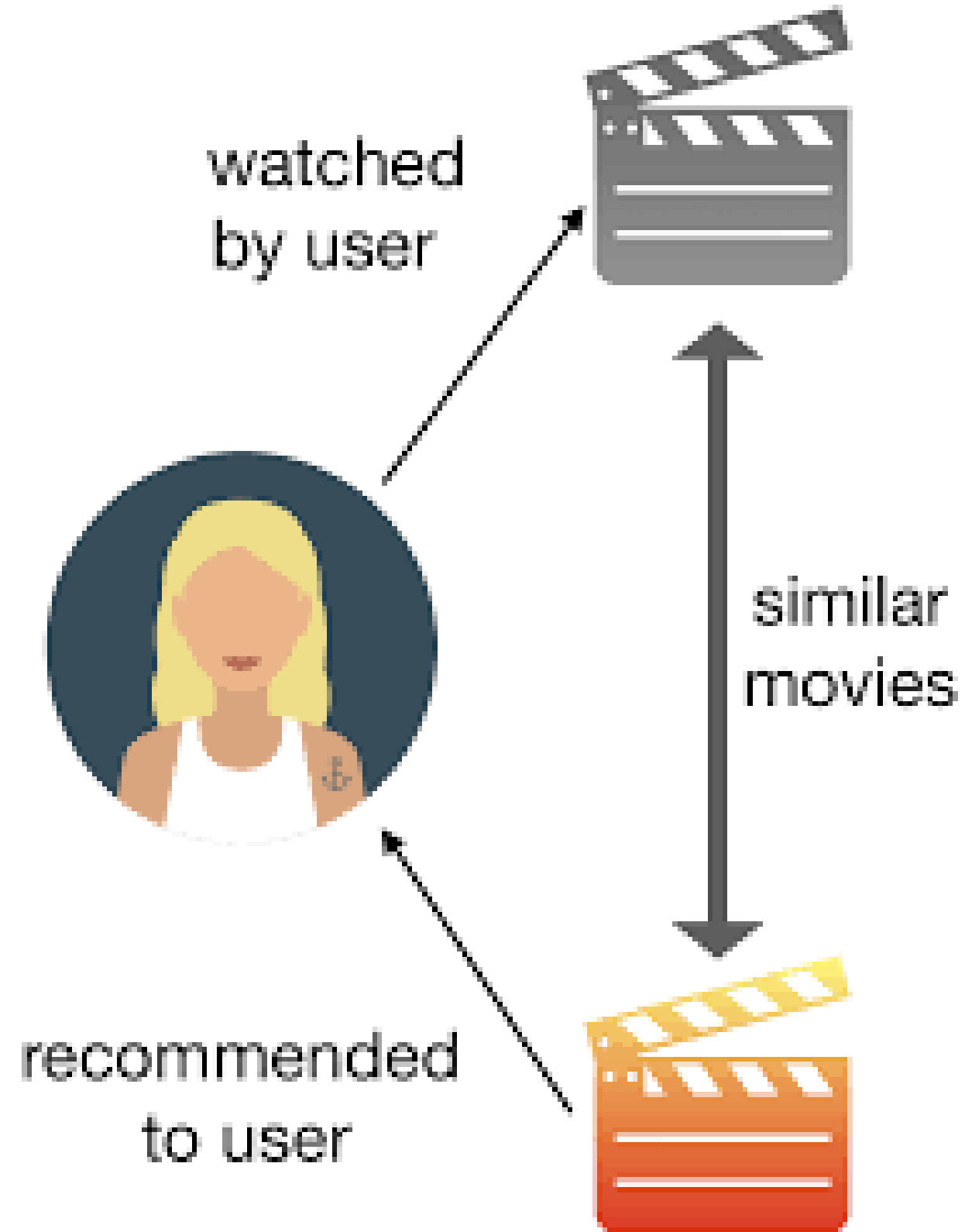
MODELS USED



BEST MODEL

- SVD with RMSE of 0.862

MODEL USED



CONTENT-BASED FILTERING

- Matrix of Pivot Table of key features
- Compute Cosine Similarity using KNN
- Make recommendations based on similarity

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



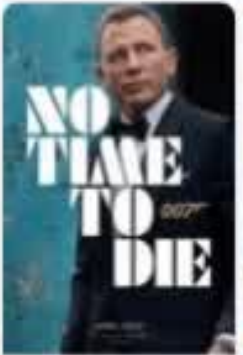
Avengers:
Endgame
2019




Da 5 Bloods
2020




Parasite
2019




No Time to ...
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



CONCLUSION

- One can conclude that developing recommender systems based on machine learning might be beneficial for services if implemented effectively, in terms of giving close predictions

RECOMMENDATION

- We recommend that taking the recommendation engine to production and integrate with the movie streaming service to help boost viewers satisfaction and revenue for our client.

Thank you!

