NETFLIX RECOMMENDATION SYSTEM

SOCIAL NETWORKS & RECOMMENDATION SYSTEMS PROJECT 13

Elżbieta Jowik, Mateusz Wójcik

AGENDA

01
PROJECT'S PURPOSE
What is the project about?

02
DATASET
The description of the dataset

03
METHODOLOGY
The description of the system

04
RESULTS
Test recommendations predicted by the algorithm

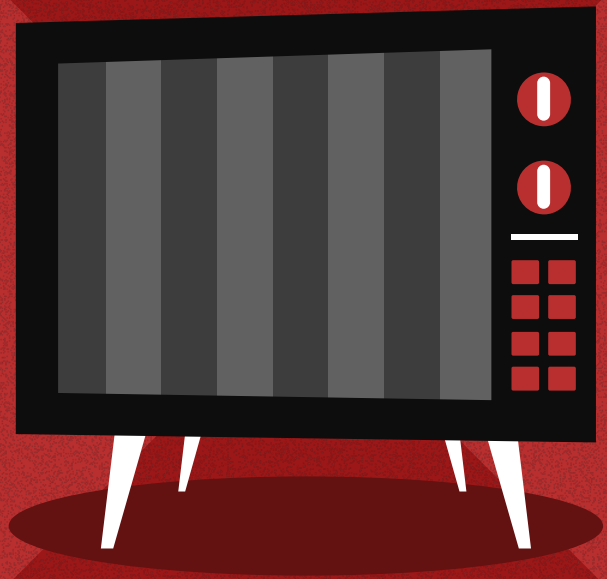
05
QUESTIONS
Questions & Answers section

PURPOSE OF THE PROJECT

The aim of the project is to implement our own recommendation system for the most popular movies among *similar* users.

The project is mostly focused on the precise implementation and testing on one of the most popular streaming platform in the world!

RS!



NETFLIX COMPETITION

Netflix organised the Netflix Prize open competition for the best algorithm that predicts users' ratings and recommends movies for the users.

The winner, BellKor's Pragmatic Chaos team, won $1,000,000.

02

DATASET

The description of the dataset

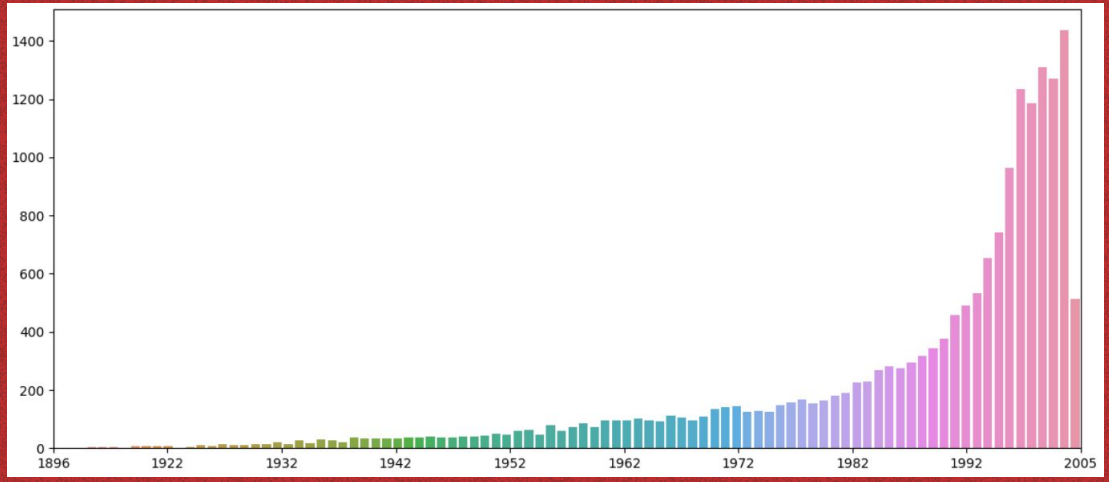
GENERAL INFORMATION

Training dataset consists of

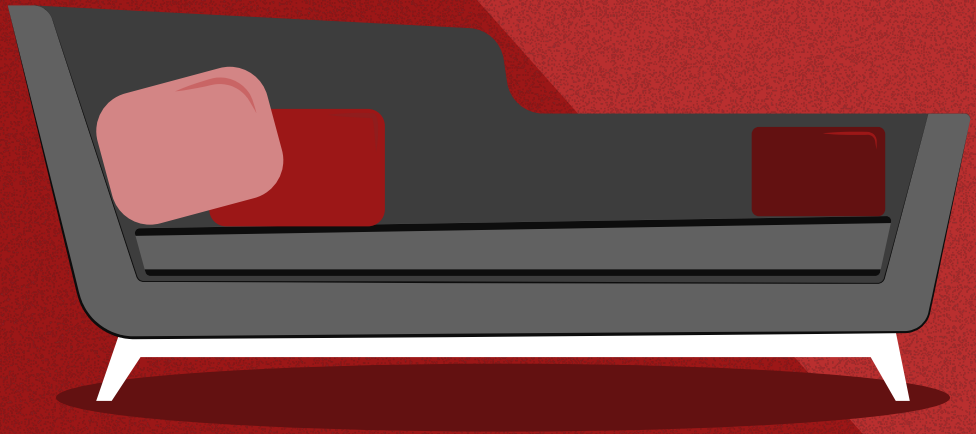
- 17 700 movies
- 480 189 users
- Customers' ratings with dates

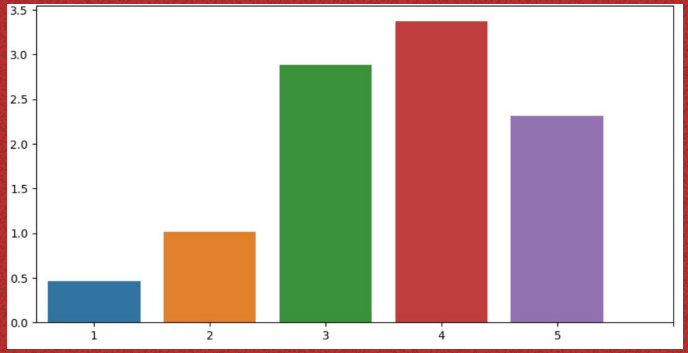
Ratings are on a five star scale – from 1 to 5.

EXPLORATORY DATA ANALYSIS



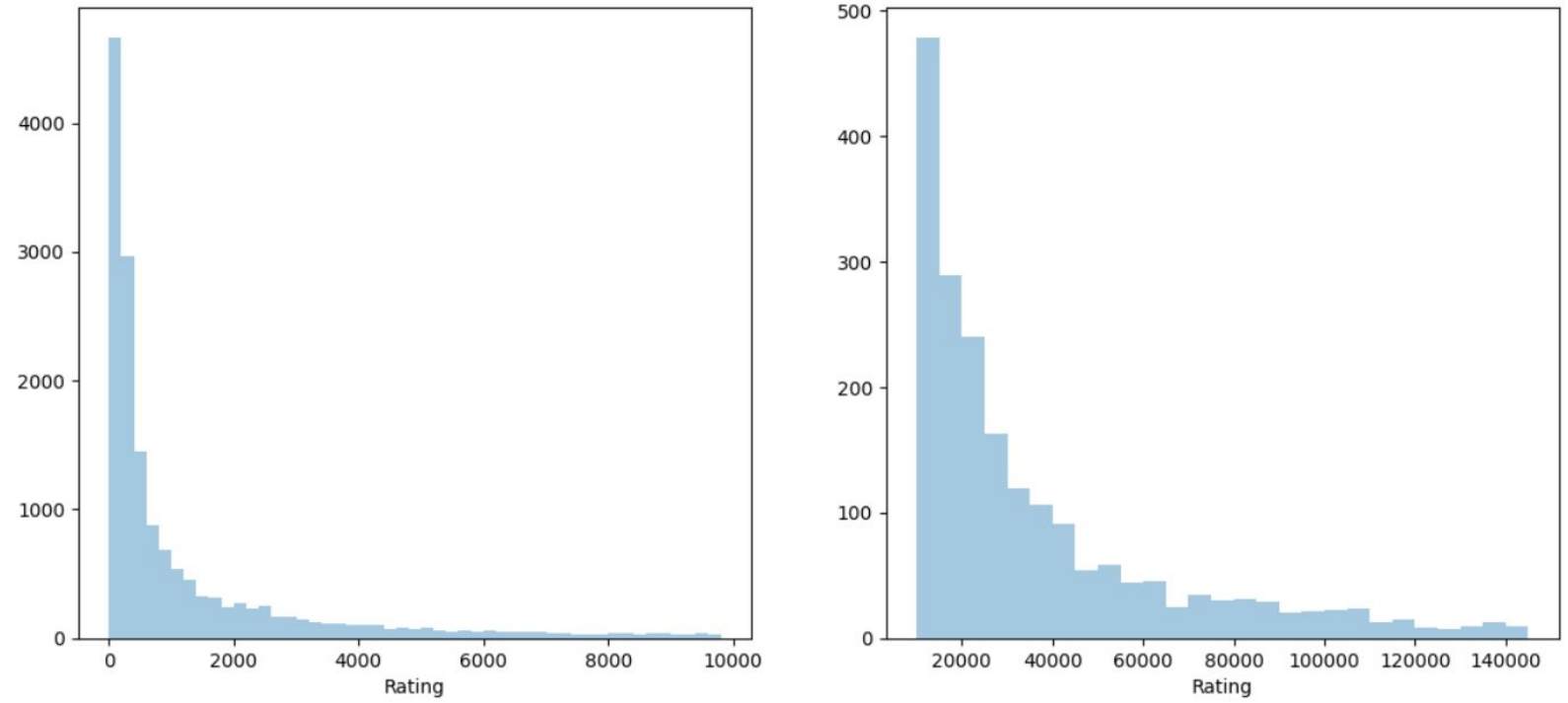
Number of movies released in a particular year



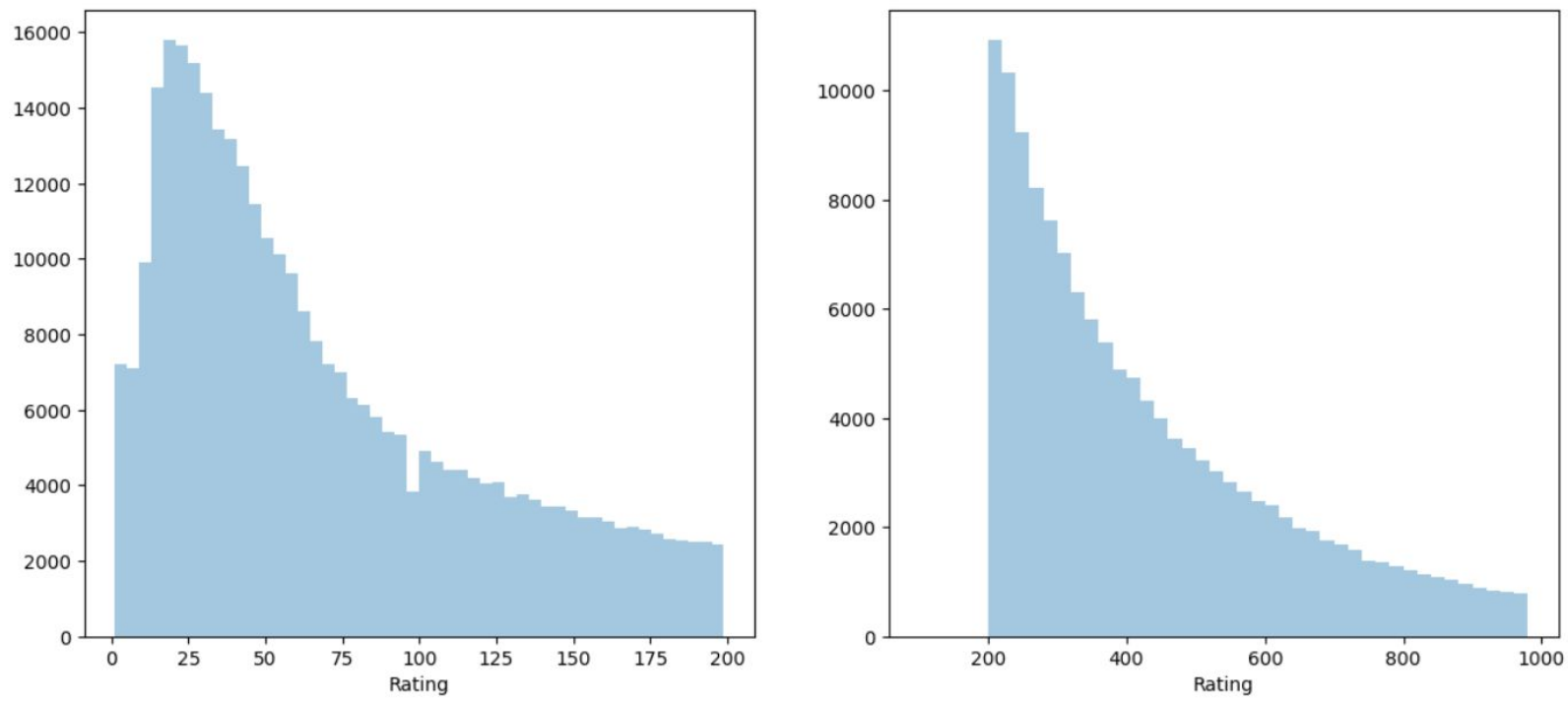


Distribution of ratings (in 10 millions)

THE NUMBER OF MOVIES WITH RATING COUNT



THE NUMBER OF USERS WITH RATING COUNT



03

METHODOLOGY

The description of the recommendation system

STEPS OF THE ALGORITHM

USERS

We compute the similarity between users identified by ***user_id*** passed to the constructor of the class.

RATINGS

Then, we try to predict the ratings of movies based on the similarities.

RATINGS' PREDICTION

A

WEIGHTED AVERAGE

The ratings are calculated using a weighted average of the ratings of movies given by the other users.

B

WEIGHTS

Weights are equal to the similarity of two users.

C

SIMILARITY

The similarity is computed based on the Pearson correlation scaled to [0, 1] interval.

DATASET ASSUMPTIONS

SUBSET?

We were able to perform calculations and exploratory data analysis on the **whole dataset**

LIMIT OF USERS?

We considered **all 17769** users for our system

04

RESULTS

The performance of our recommendation system

EXAMPLE OUTPUT OF THE SYSTEM

2635437	4.779132	4.0
479924	4.697763	5.0
1686060	4.545448	5.0
2207774	4.493986	5.0
1554712	4.421567	5.0
...	...	...
209573	2.361964	4.0
305344	1.950932	1.0
1850615	1.822477	2.0
1227322	1.735407	4.0
1567202	1.646744	2.0

METRICS

0.1756
Mean Absolute Error (MAE)

0. 7762
Mean Squared Error (MSE)

42.18%
Accuracy

THANKS!

Do you have any questions?