SPRINT 3

LETTERBOXD MOVIE RECOMMENDER

Cole Beevor- Potts

Problem

FINDING A MOVIE

That isn't terrible and doesn't take 45 minutes to find.

109 HOURS!

Average Netflix user browsers for 17.8 minutes before choosing a show.

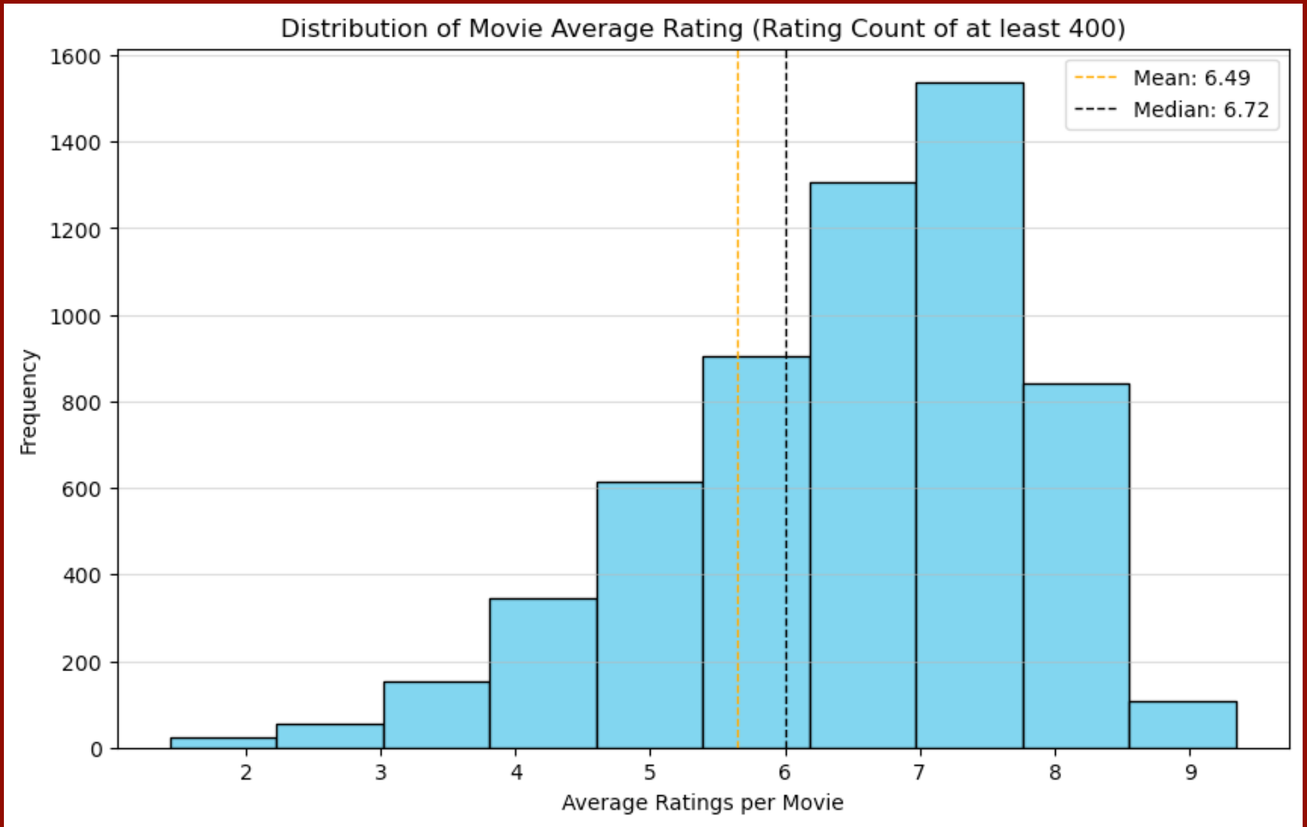
Solution

MOVIE RECCOMENDER

A personalized movie recommender that puts your tastes first.

4.5 DAYS BACK PER YEAR!

My Goal



- Average score for a movie is 6.49 out of 10

- Make your decision :
  - **Easier**
  - **Faster**
  - **More Enjoyable**

Data & Processing

THE DATA

MOVIE DATA
280k movies
19 features

RATING DATA
11 million ratings
7477 users

APPROACH

CONTENT BASED RECCOMENDATION
Movie description

COLLABORATIVE FILTERING
Using movie reviews and users

PRE-PROCESSING & METHODS

TF_IDF VECTORIZER
Movie description

HUGGING FACE TRANSFORMER
Using BERT

MATRIX FACTORIZATION
FunkSVD

PART 1:

Content Based Reccomendation

MODEL

NEAREST NEIGHBOURS

- Returns k nearest neighbours
- Finds the closest vector

APPROACH 1

TF-IDF VECTORIZER

- Counts the occurrences of words
- Higher weight on less common words
- Picking up on the wrong things

APPROACH 2

HUGGING FACE TRANSFORMERS

- BERT - Bidirectional Encoder Representations from Transformer
- GPT - Generative Pre-trained Transformer
- **Smart & contextual**

PART 2:

Collaborative Filtering

$$R = U \cdot M$$

$$R = \begin{bmatrix} 4 & 1 & 1 & 5 \\ 5 & 5 & 2 & 1 \\ 3 & 5 & 4 & 3 \end{bmatrix}$$

$$U = \begin{bmatrix} 0.029 & 2.898 & 0.46 \\ 2.33 & 0.61 & 0.0 \\ 0.561 & 0.0 & 1.87 \end{bmatrix}$$

$$M = \begin{bmatrix} 1.78 & 2.13 & 0.75 & 0.0 \\ 1.36 & 0.0 & 0.37 & 1.63 \\ 0.0 & 2.03 & 1.91 & 0.53 \end{bmatrix}$$

- MATRIX FACTORIZATION USING FUNK SVD
- PREDICT USER RATINGS
- ROOT MEAN SQUARE ERROR OF 1.325

Reccomender Demonstration

Next Steps

- **COMBING MODELS**
  - FIND SIMILAR MOVIES BASED ON CONTENT
  - RECOMMEND THE MOVIE THE USER WOULD PREDICT THE HIGHEST

- **LETTERBOXD SCRAPER**
  - TAKE IN AND FIT A USERS DATA FROM LETTERBOXD