# Movie Recommender System

Priscilla
26-01-2023

# Content

***SECTION:*** **01**

## Introduction

Background
Problem Statement

***SECTION:*** **02**

## EDA

Genres
Ratings
Unique Words

***SECTION:*** **03**

## Recommender Systems

Basic
Content-Based
Collaborative

***SECTION:*** **04**

## Conclusion

Advantages vs Disadvantages
Recommendations

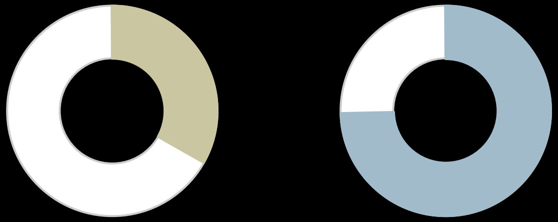
ROLL: **01**

TAKE: **Introduction**

# Recommender Systems

| Movie Streaming Service | E-Commerce |
|---|---|
| Netflix | Amazon |

# Benefits to Businesses/Consumer



**35%** Amazon **75%** Netflix

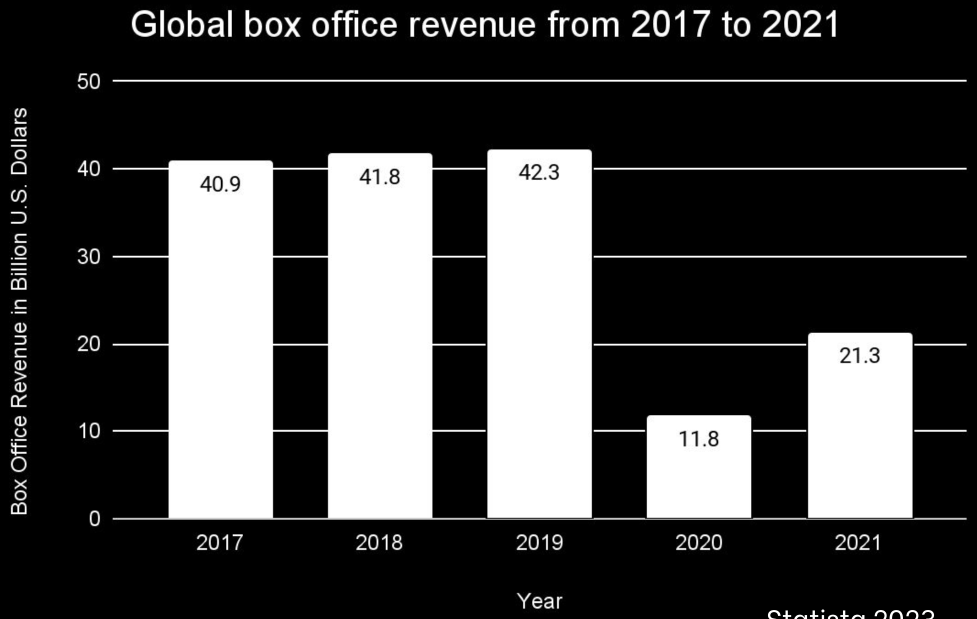
McKinsey

**35%** of what consumers purchase on Amazon and **75%** of what they watch on Netflix comes from product recommender systems

# Benefits to Consumers/Businesses

- Customer Satisfaction
- Personalization/Discovery
- Revenue

# Film Industry




Statista 2023

- Steadily **increasing** over the years before COVID-19 in 2020
- However, revenue is in an **upward trend** since 2020

# Consumers

- Fast
- Less Effort

# Problem Statement

The aim of this project is to build various types of movie recommender systems and evaluate these systems based on their performance.

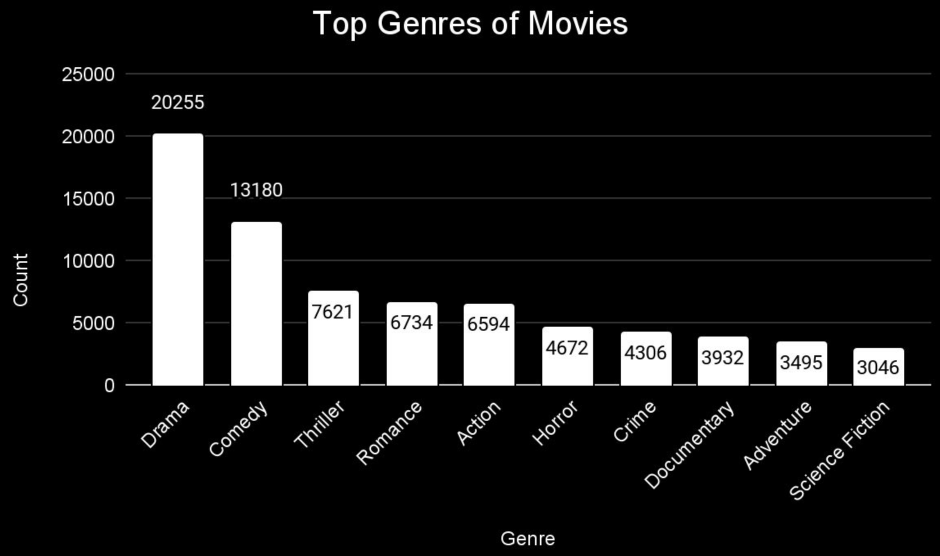
ROLL: **02**

TAKE: **EDA**

# Dataset

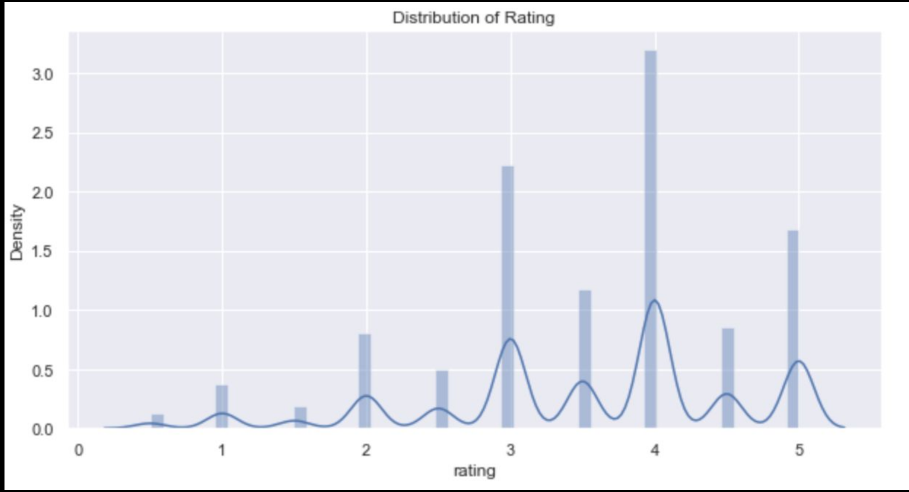
<table>
<tr><td>Metadata</td><td rowspan="3">Metadata of all 45,000 movies released on or before July 2017</td></tr>
<tr><td>Keywords</td></tr>
<tr><td>Credits</td></tr>
<tr><td>Ratings</td><td>26 million ratings from 270,000 users for all 45,000 movies</td></tr>
</table>

# Genre



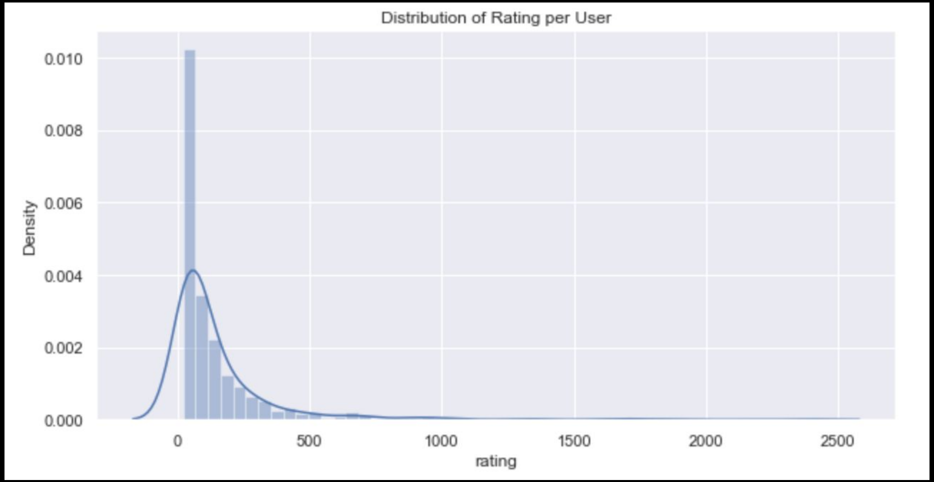

- Top genres of movies in the dataset are **Drama**, **Comedy** and **Thriller**

# Ratings



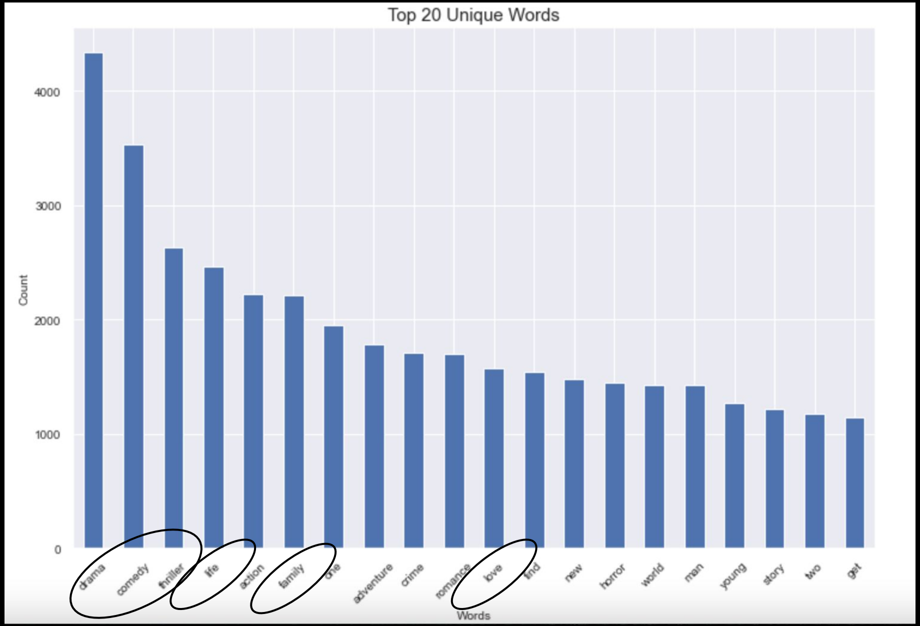

- Majority of user gave the movies a **3-4** out of 5 rating

# Ratings per User




- Each user rates a minimum of **20** movies and up to a maximum of **2391** movies.

# Top 20 Unique Words




- Using the words from movie title, overview, tagline, genres, keywords and cast
- Top words are **Drama**, **Comedy** and **Thriller** which coincides with the top 3 genres of movies
- Movies are based on the themes of **Life**, **Family** and **Love**.

ROLL: **03**

TAKE:

# Recommender Systems

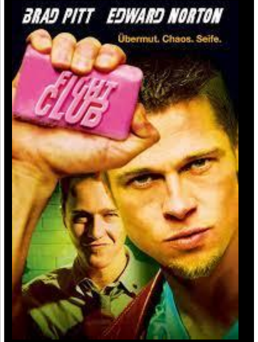
# Basic Recommender System

IMDB Weighted Rating Formula

**Weighted Rating (WR) = v/(v+m.R) + m/(v+m.C)**

where,
v is the number of votes for the movie
m is  the minimum number of votes required to be listed in the charts
R is the average rating of the movie
C is the mean vote across the whole report

# Top 5 movies - WR

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **Movie** |   |   |   |   |   |
| **WR** | 8.3 | 8.0 | 8.0 | 8.0 | 8.0 |

# Top 5 movies - WR and Genre

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **Movie** |  |  |  |  |  |
| **WR** | 7.9 | 7.9 | 7.9 | 7.8 | 7.8 |

# Content-Based Filtering

***STEP:*** **01**

**Select the Categories of interest**

Title, Overview, Tagline, Genres, Keywords, Cast

***STEP:*** **02**

**Merge all columns into one column**

tags

***STEP:*** **03**

**Clean Text**

Remove punctuations, stopwords and lemmantization

# Content-Based Filtering

***STEP:*** **04**

**Apply TF-IDF**

n-gram = 1

***STEP:*** **05**

**Apply Consine Similarities**

***STEP:*** **06**

**Find similar movies based on movie attributes**

# Content-Based Filtering

**Minions**

| | |
|---|---|
| 1 | Mower Minions (0.413) |
| 2 | Minions: Orientation Day (0.362) |
| 3 | Banana (0.273) |
| 4 | Despicable Me 2 (0.251) |
| 5 | Despicable Me (0.215) |

**The Dark Knight**

| | |
|---|---|
| 1 | The Dark Knight Rises (0.307) |
| 2 | Batman Begins (0.273) |
| 3 | Batman Returns (0.270) |
| 4 | Batman Forever (0.257) |
| 5 | Batman: Under the Red Hood (0.229) |

# Collaborative Filtering

## Collaborative Filtering

Filters movies a user might like based
on the ratings given by similar users

### Memory-Based

User-Based
Item-Based

### Model-Based

Matrix Factorization
based on SVD and NMF

# Memory-Based Collaborative Filtering

## Surprise Library

### User-Based

KNNBasic Algorithm
Consine Similarities
User-Based: True

### Item-Based

KNNBasic Algorithm
Consine Similarities
User-Based: False

# Model-Based Collaborative Filtering

## Surprise Library

### SVD

RMSE: 0.87

### NMF

RMSE: 0.95

# User 671

**Mystery/Thriller**

**Drama/Comedy**

**Action/Sci-Fi**

| |
|---|
| The Poseidon Adventure |
| The Million Dollar Hotel |
| Terminator 3 |
| The Searchers |
| Boogie Nights |
| Jacob's Ladder |

# User 671

| User-Based |
| --- |
| Blown Away |
| Rio Bravo |
| The Celebration |
| Stalag |
| Gentlemen Prefer Blondes |

| Item-Based |
| --- |
| 2046 |
| The Protector |
| Everything is Illuminated |
| K-PAX |
| The Silence of the Lambs |

# User 671

| SVD |
|---|
| Nell |
| Flags of Our Fathers |
| Fools Rush In |
| Shriek If You Know What I Did Last Friday the Thirteenth |
| Sleepless in Seattle |

| NMF |
|---|
| Gentlemen Prefer Blondes |
| Once Upon a Time in Mexico |
| Ninotchka |
| Speed Racer |
| Backdraft |

ROLL: **04**

TAKE: **Conclusion**

# Conclusion

| | Advantage | Disadvantage |
|---|---|---|
| **Basic** | • Simple<br>• Easy to build | • Not unique to user preference<br>• Lack of personalization |
| **Content-Based Filtering** | • Recommendations are specific to user, do not require other user data, can scale to large number of users | • Cold Start<br>• Make recommendations based on existing interest of user |

# Conclusion

| | Advantage | Disadvantage |
|---|---|---|
| **Collaborative Filtering** | • All users are taken into consideration and people with similar tastes and preference are used to suggest new movies to the primary user | • Sparsity<br>• Scalability |

# Recommendations

1. Explore Deep Neural Networks for Collaborative Filtering
2. Explore Hybrid Content-Based and Collaborative Filtering Recommender Systems

# THANKS!

ROLL:

**DO YOU HAVE ANY QUESTIONS?**

youremail@freepik.com
+91  620 421 838
yourcompany.com

TAKE:



SCENE: