Movie Recommender System

Jack Frantz

Problem Statement

- Movie Database Startup Application
  - Stores the user's movie watching history and ratings
- Use data from users to help recommend new content
- Build a recommender system that will:
  - Take in movie ratings and a search from users
  - Return movies that similar users enjoyed

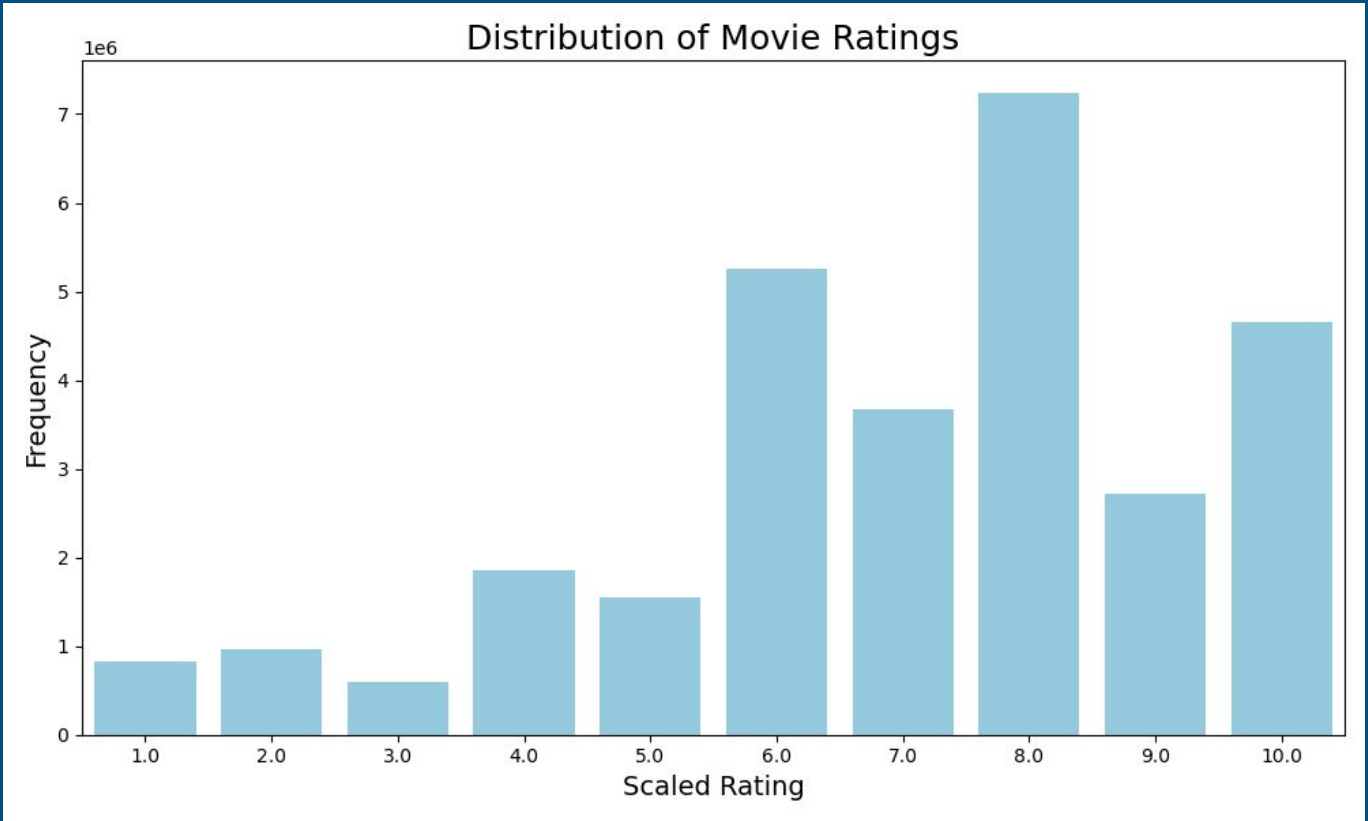
Data

- Reviews from active movie databases
  - MovieLens
  - IMDb
- Movie features from IMDb Non-Commercial Datasets
- Will use existing these existing databases as blueprint

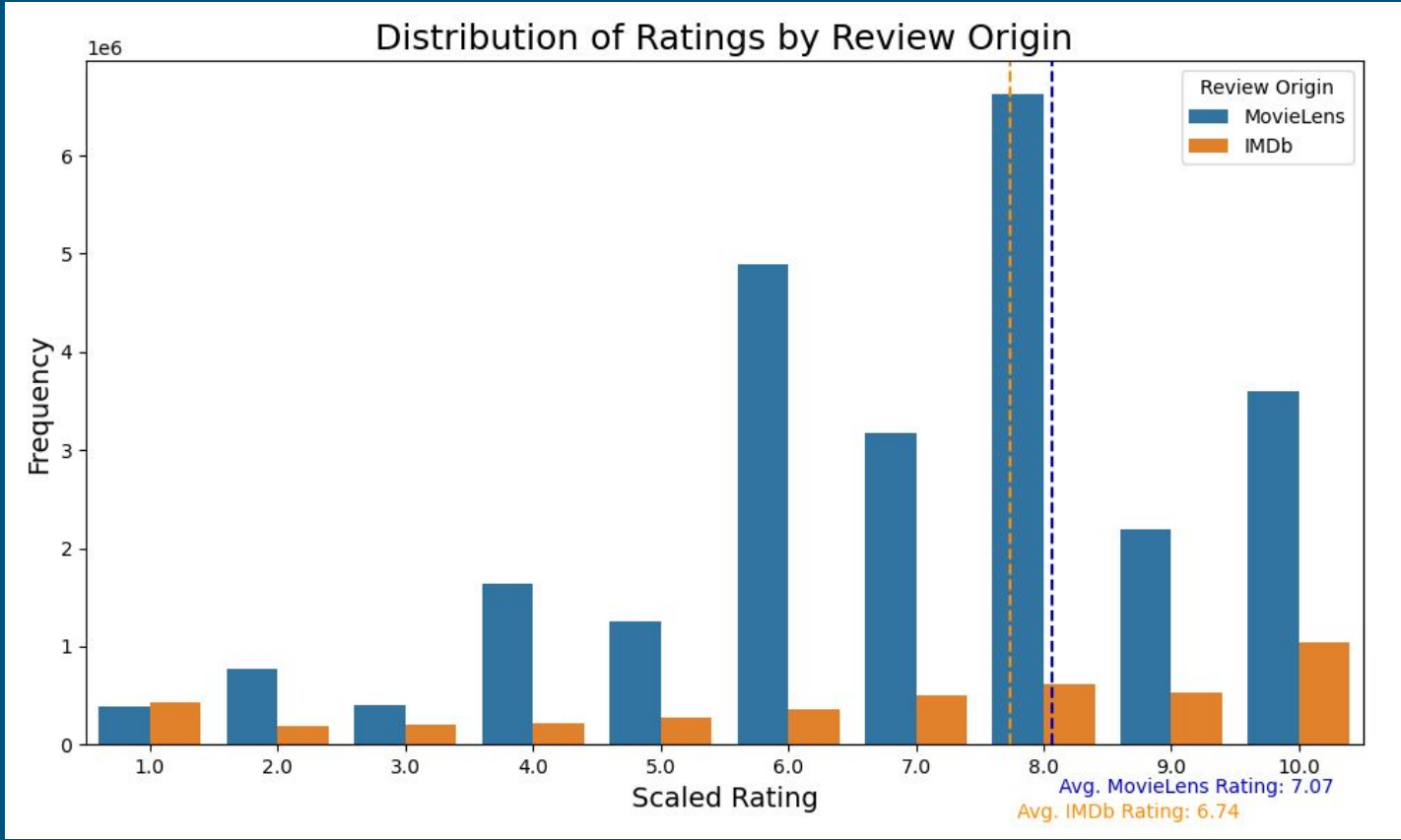
Review Data

- IMDb: Scale of 1-10
  - 4,669,820 Ratings
  - 1,499,238 Users
  - 351,109 Movies
- MovieLens: Scale of 0-5
  - 25,000,000 Reviews
  - 162,000 Users
  - 62,000 Movies

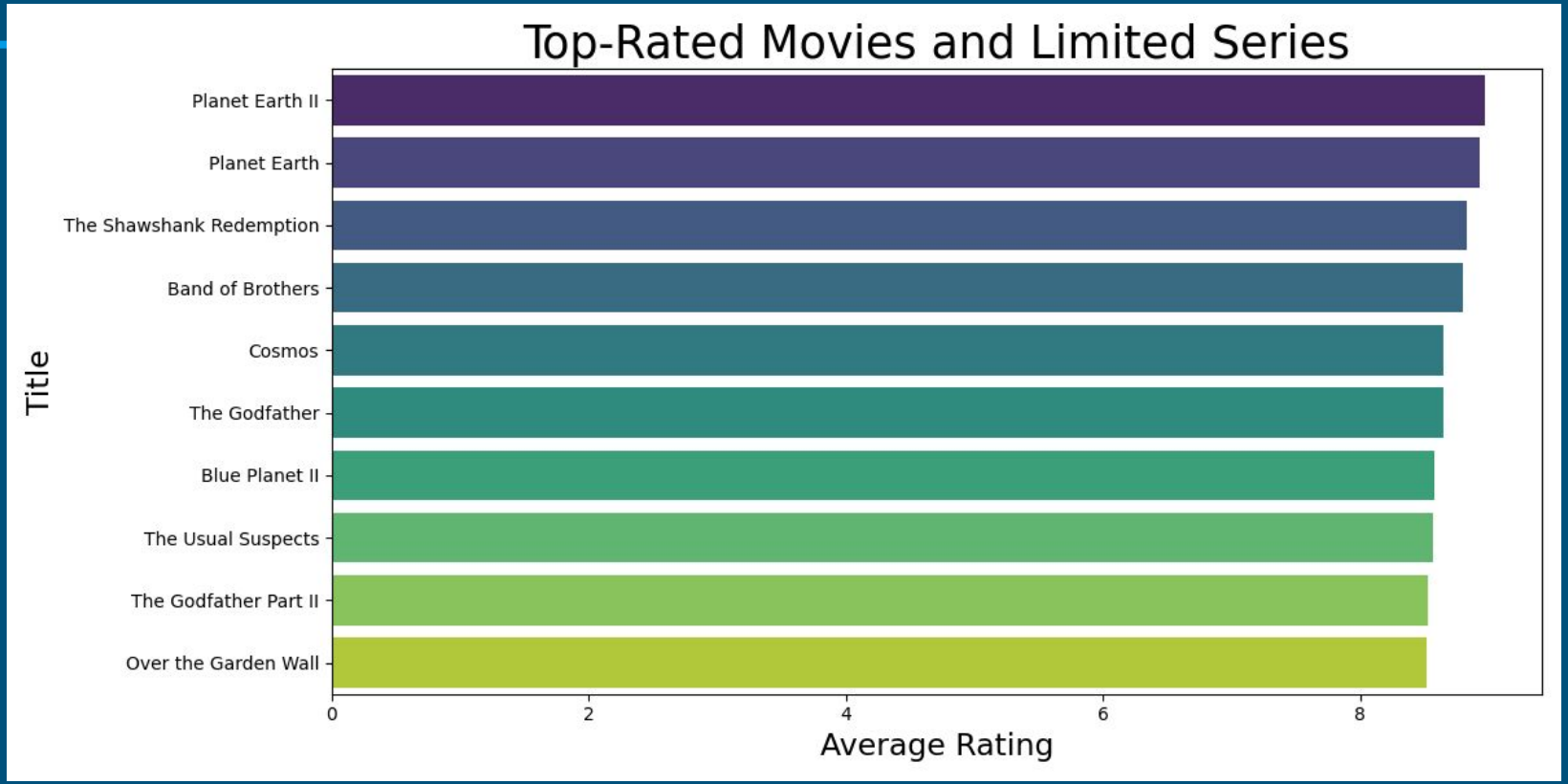
Distribution of Movie Ratings



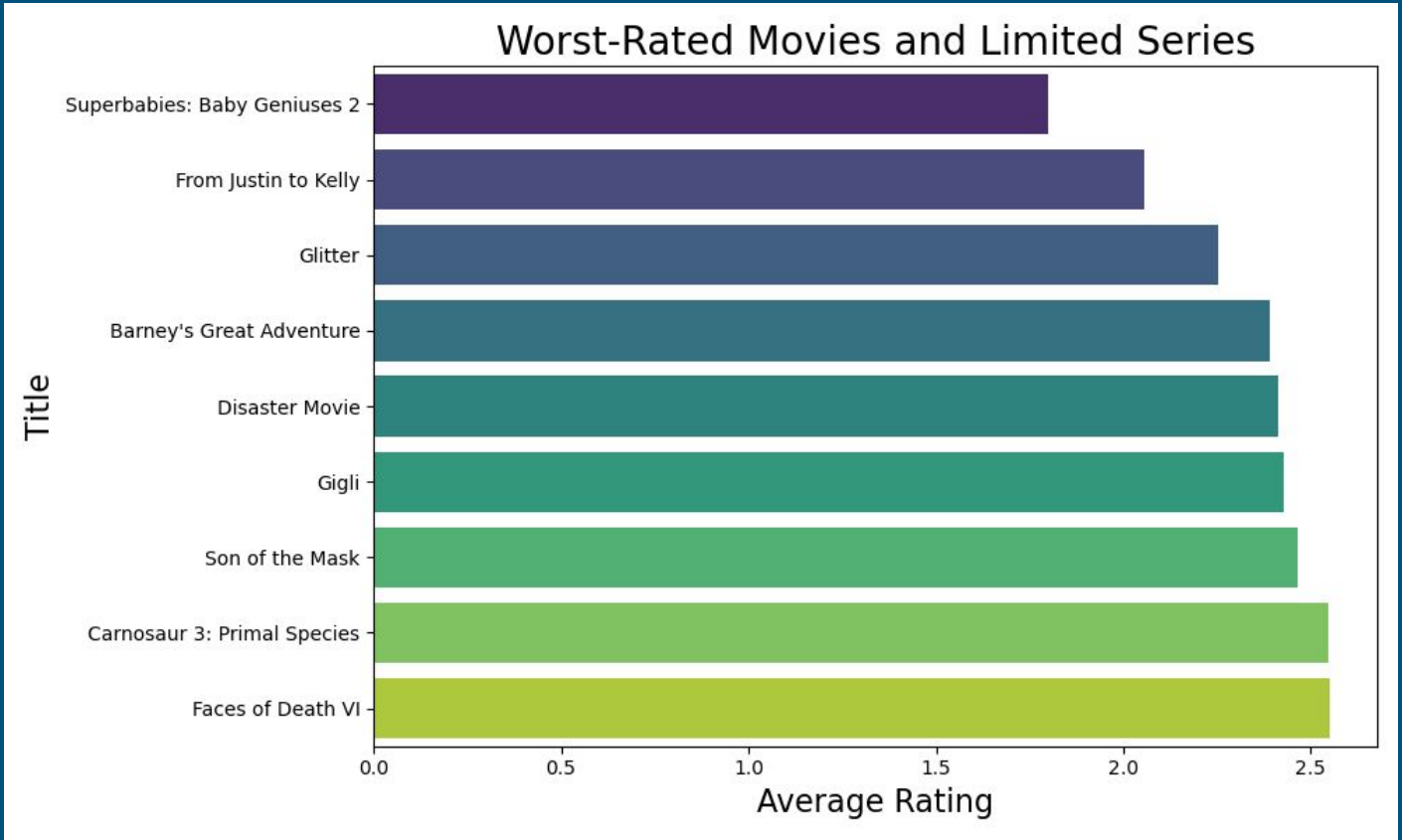
Distribution of Ratings by Source



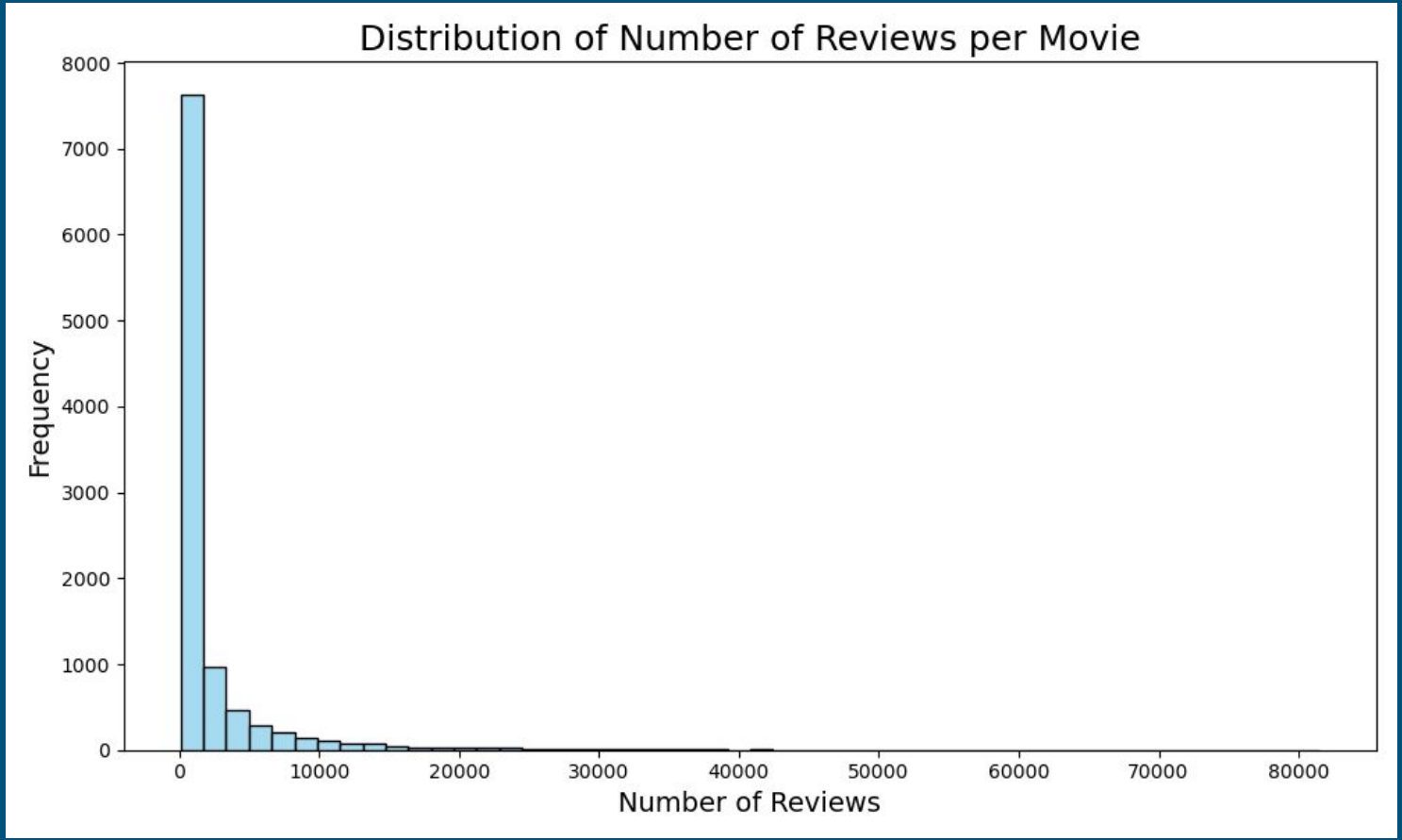
Top Rated Movies



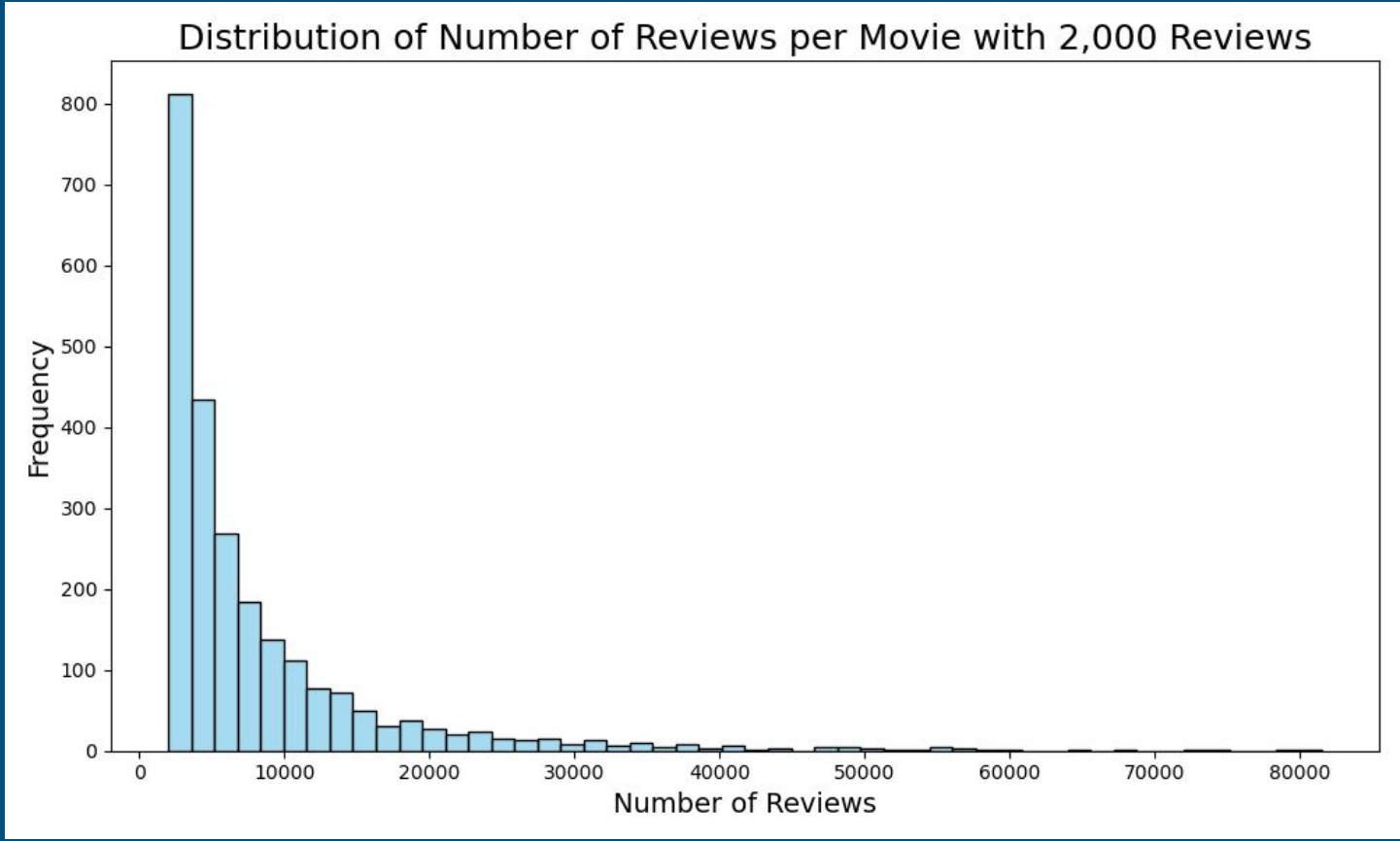
Worst Rated Movies



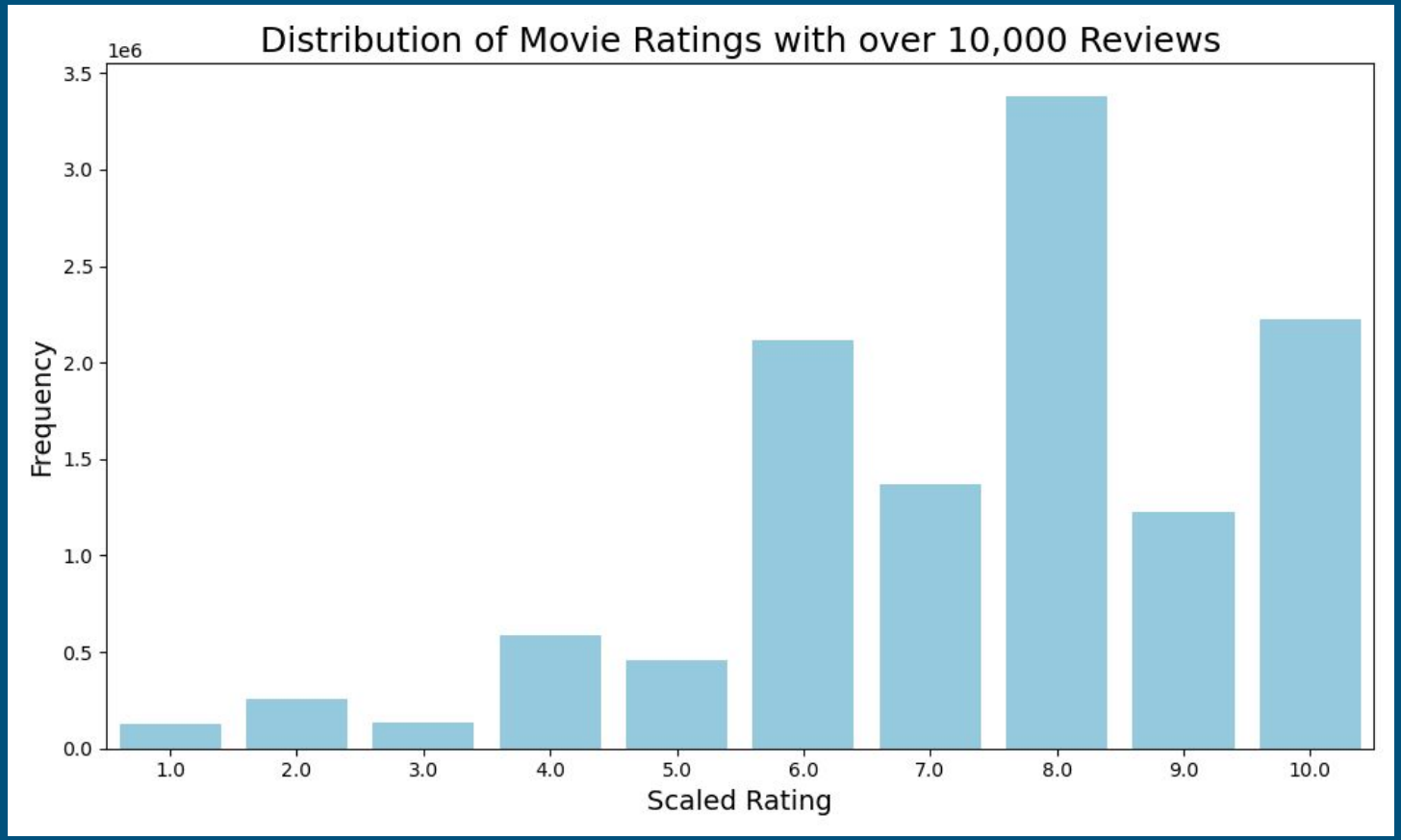
Reviews per Movie



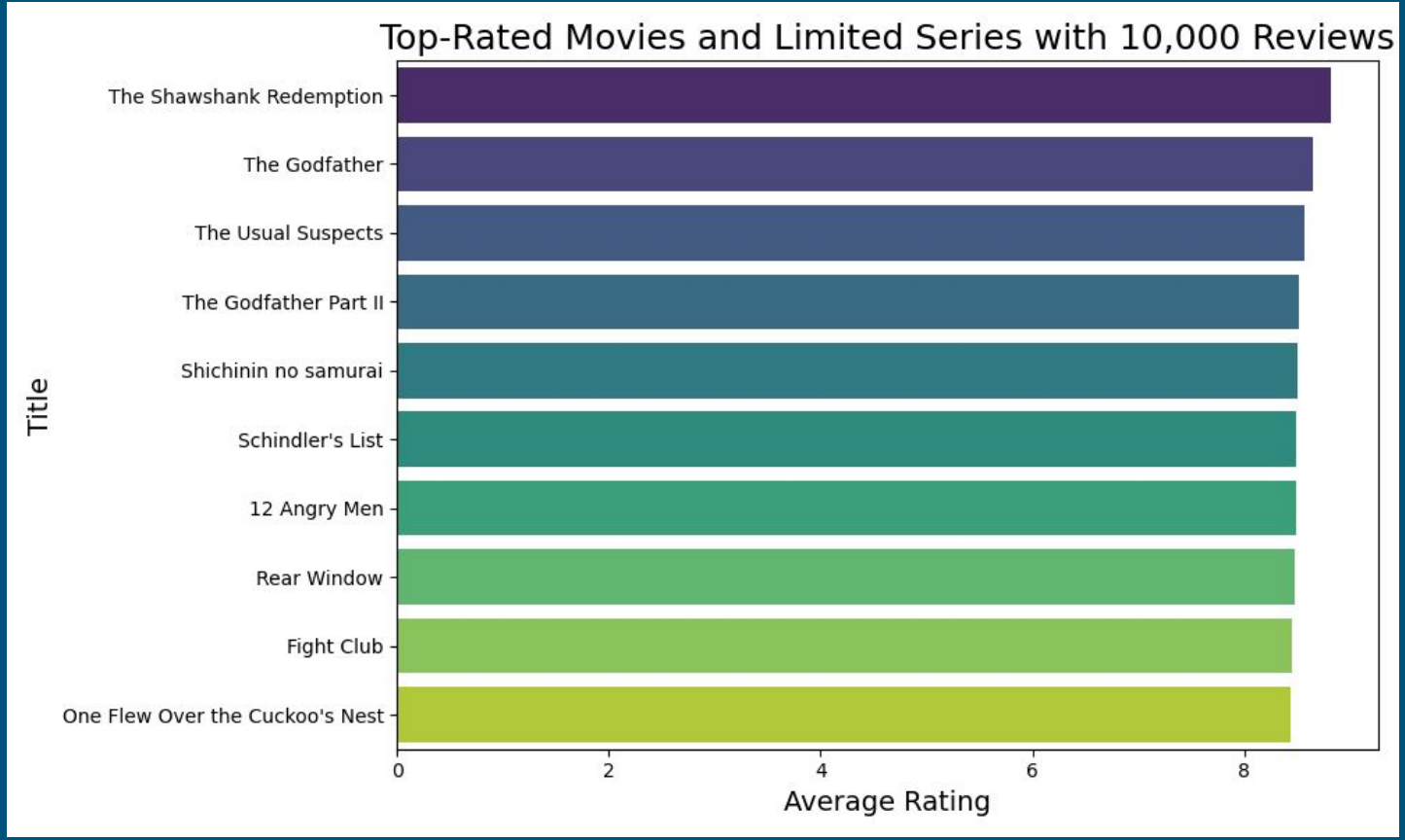
Reviews per Movie with Minimum 2k Reviews



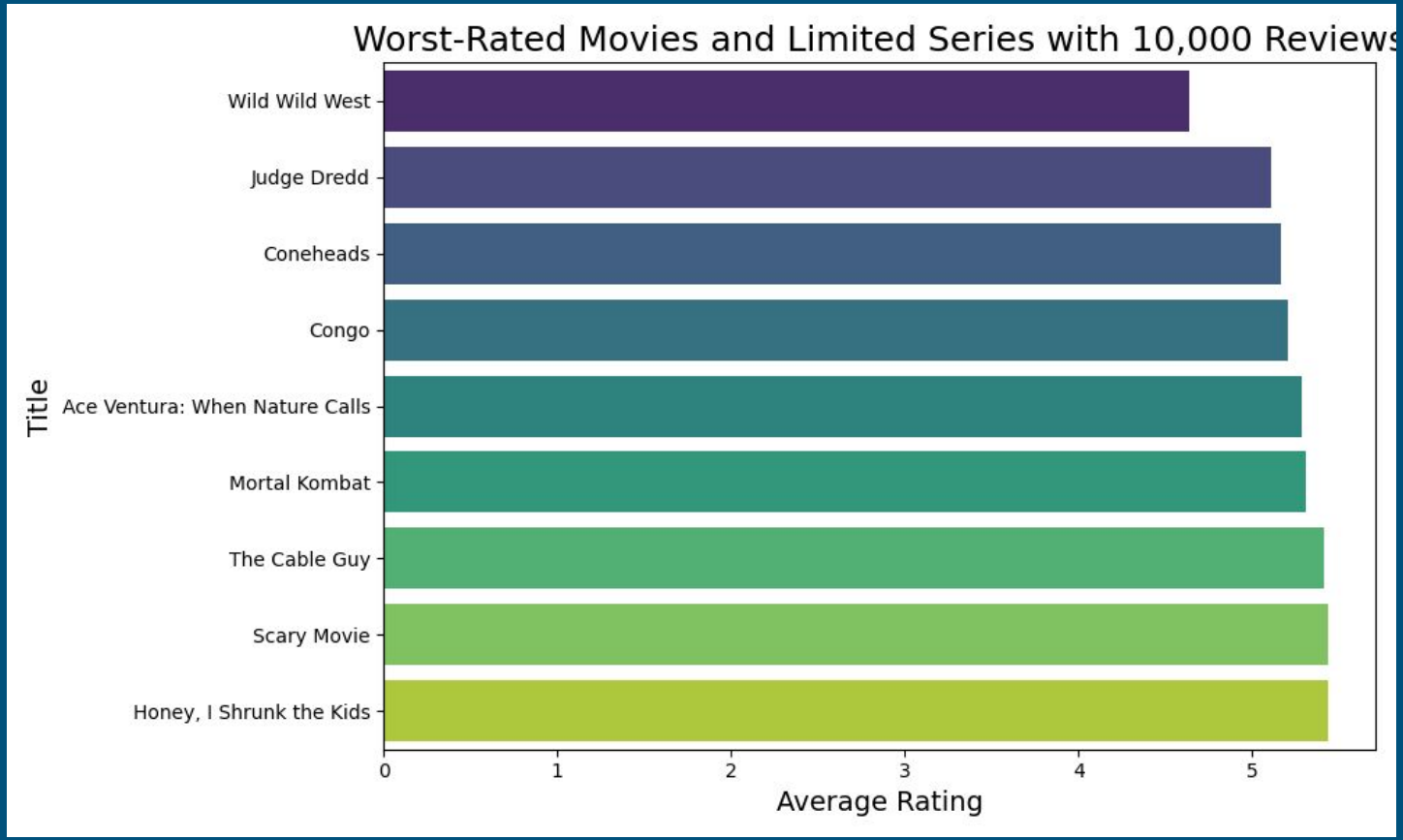
Distribution of Ratings with 10,000 Reviews



Top Rated Movies with 10,000 Reviews



Worst Rated Movies with 10,000 Reviews



Parameters for Recommender System

- MovieLens Data only
  - Fewer unnecessary features with insufficient data
- Only half of 25M reviews that were available
  - Computing power was a limiting factor
- Cosine Similarities

```
Movies like: Toy Story
====================
title
Star Wars                                          0.564555
Toy Story 2                                        0.564336
Back to the Future                                 0.550556
Forrest Gump                                       0.546750
Jurassic Park                                      0.541597
Independence Day                                   0.537916
Star Wars: Episode VI - Return of the Jedi         0.537413
The Lion King                                      0.528811
Aladdin                                            0.525246
Star Wars: Episode V - The Empire Strikes Back     0.513098
```

Summary of Analysis

- The amount of data in the model increases the accuracy of its recommendations
- Movies with fewer reviews can skew results if included
- The number of unique users and movies increased demand for computing power most

Recommendations for Movie Database Startup

- Begin Collecting User Data
- Build an application for users to store
- Minimum threshold for movies included in recommender

Next Steps

- Create Streamlit with Recommender System
  - Users can search for their own recommendations
- Access more computing power
  - Cloud Computing
  - GPU

Questions?