Movie Recommendation System

Module 4 V2.1 Project
Stacy Shingleton
May 15, 2020

What are Recommendation Systems?

Unpersonalized: Overall popular items

Personalized: Recommended items per user

Item based similarities

User based similarities

What are Recommendation Systems?

Who uses them? Podcasts, social media applications, media streaming services, online shopping, etc.

What are they recommending? Movies, television series, songs, consumer goods, podcast channels, etc.
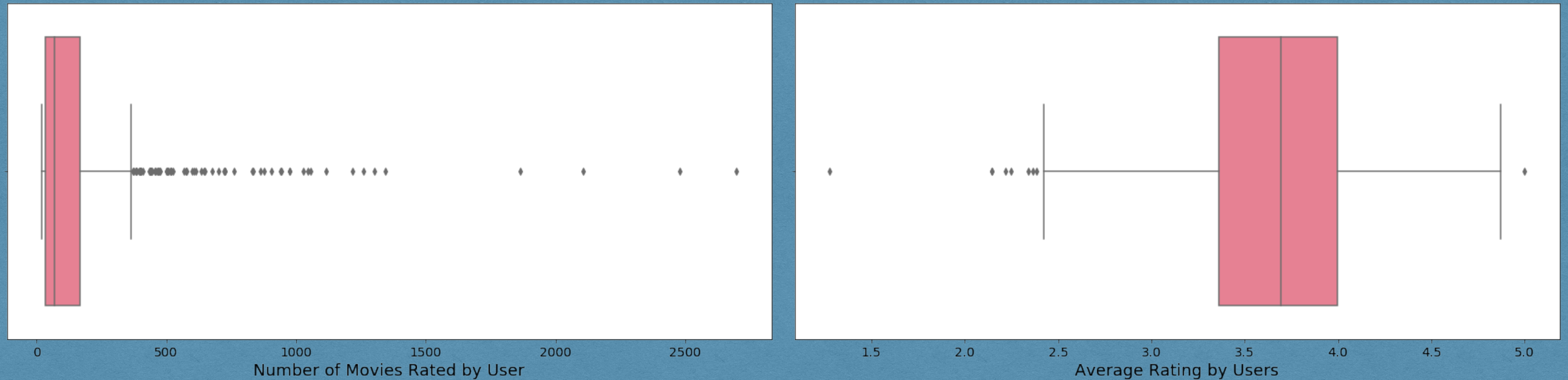
Objective

Provide movie recommendations for established or new users from a dataset of nearly 10,000 movies

The Dataset

- **GroupLens: MovieLens: Small Dataset**

  - **Ratings Dataset:**

    - Contains user ID, movie ID, rating

    - 100,836 ratings from 610 unique users for 9,724 unique movies

  - **Movie Dataset:**

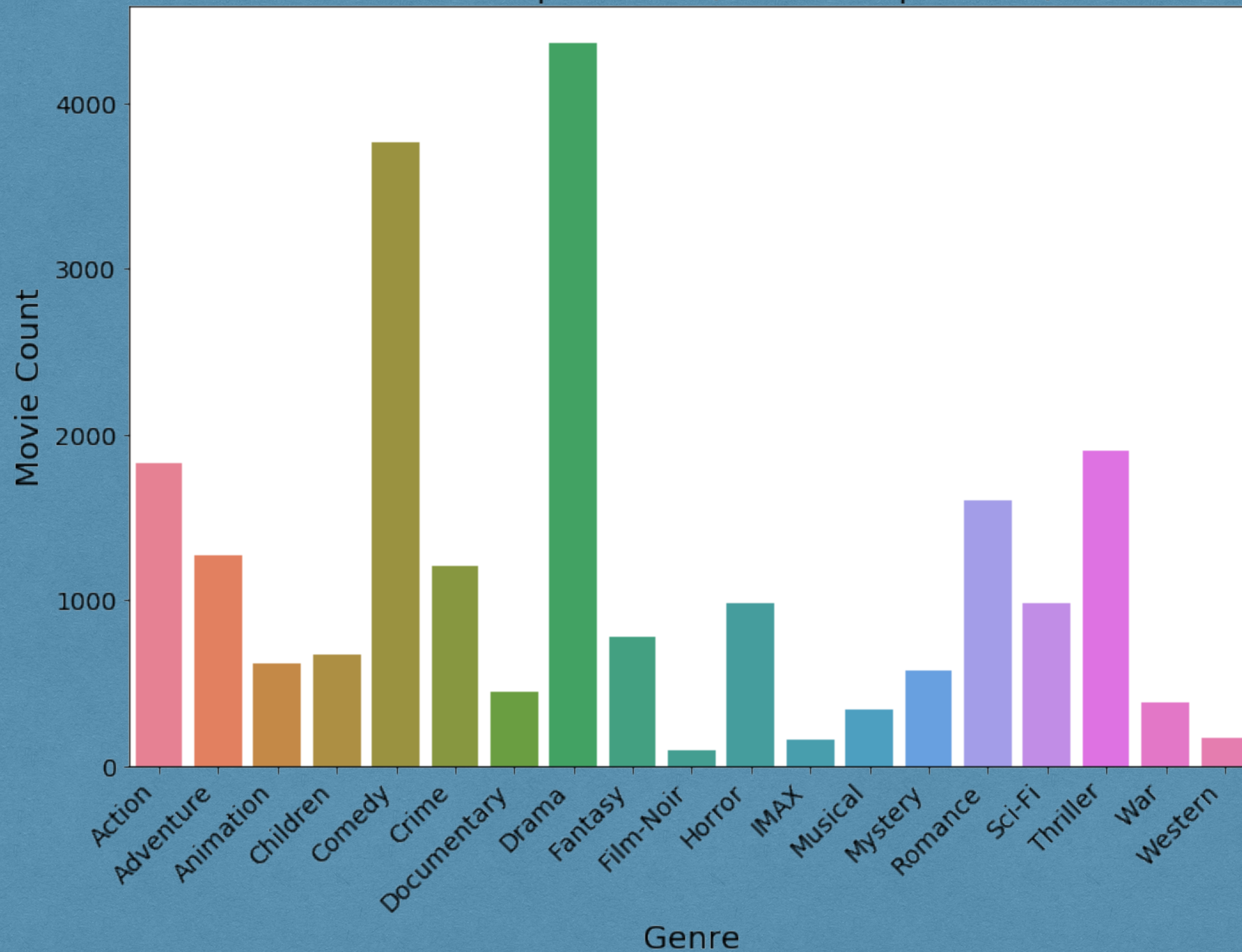   - Contains movie ID, title, genres

    - 9,742 unique movies

Filtering the Data

The ratings dataset was filtered by normalizing the number of movies rated by user, and the average rating by user. These filtered userIDs narrowed the ratings dataset from ~100k to ~30k.
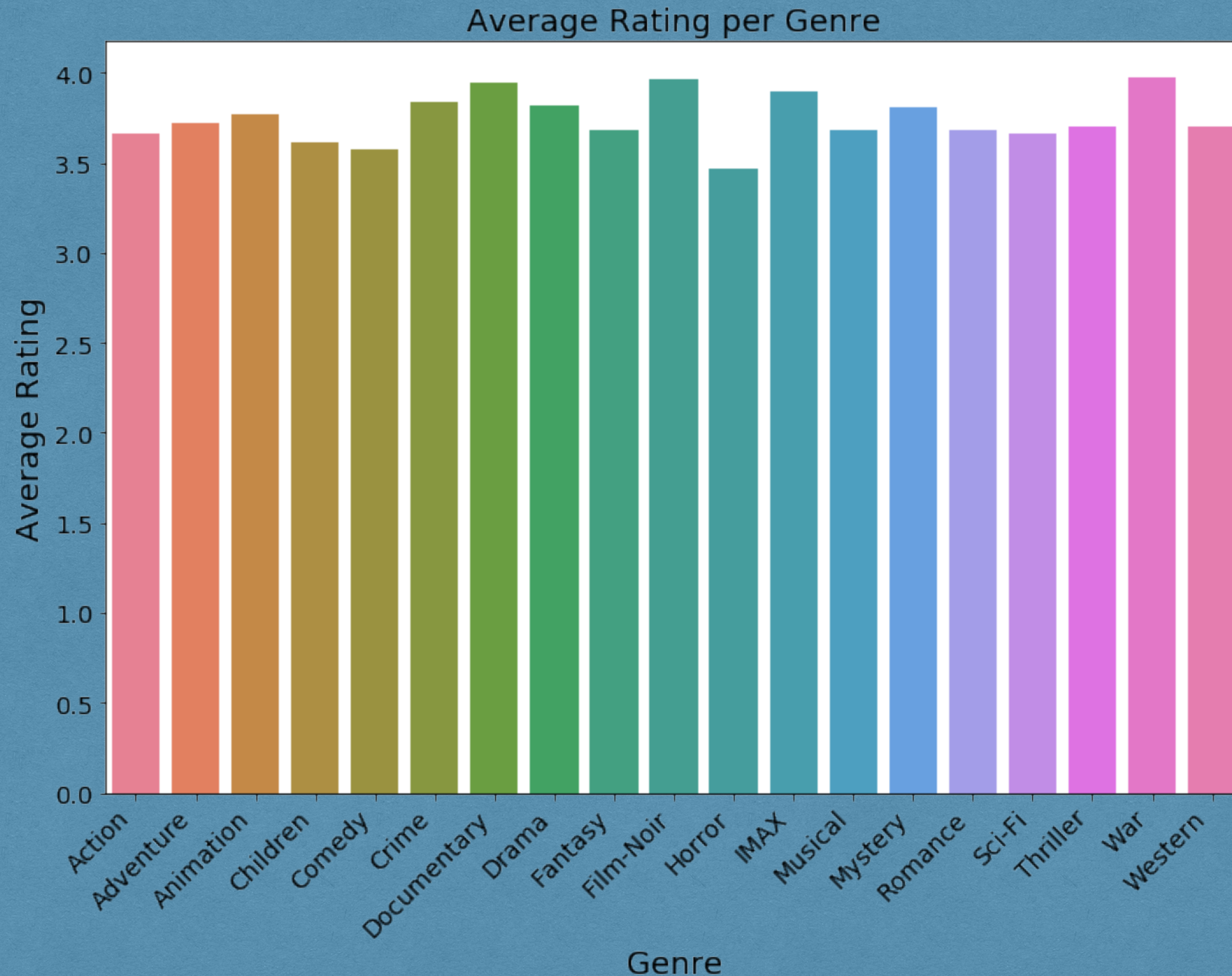
Visualizing the Data
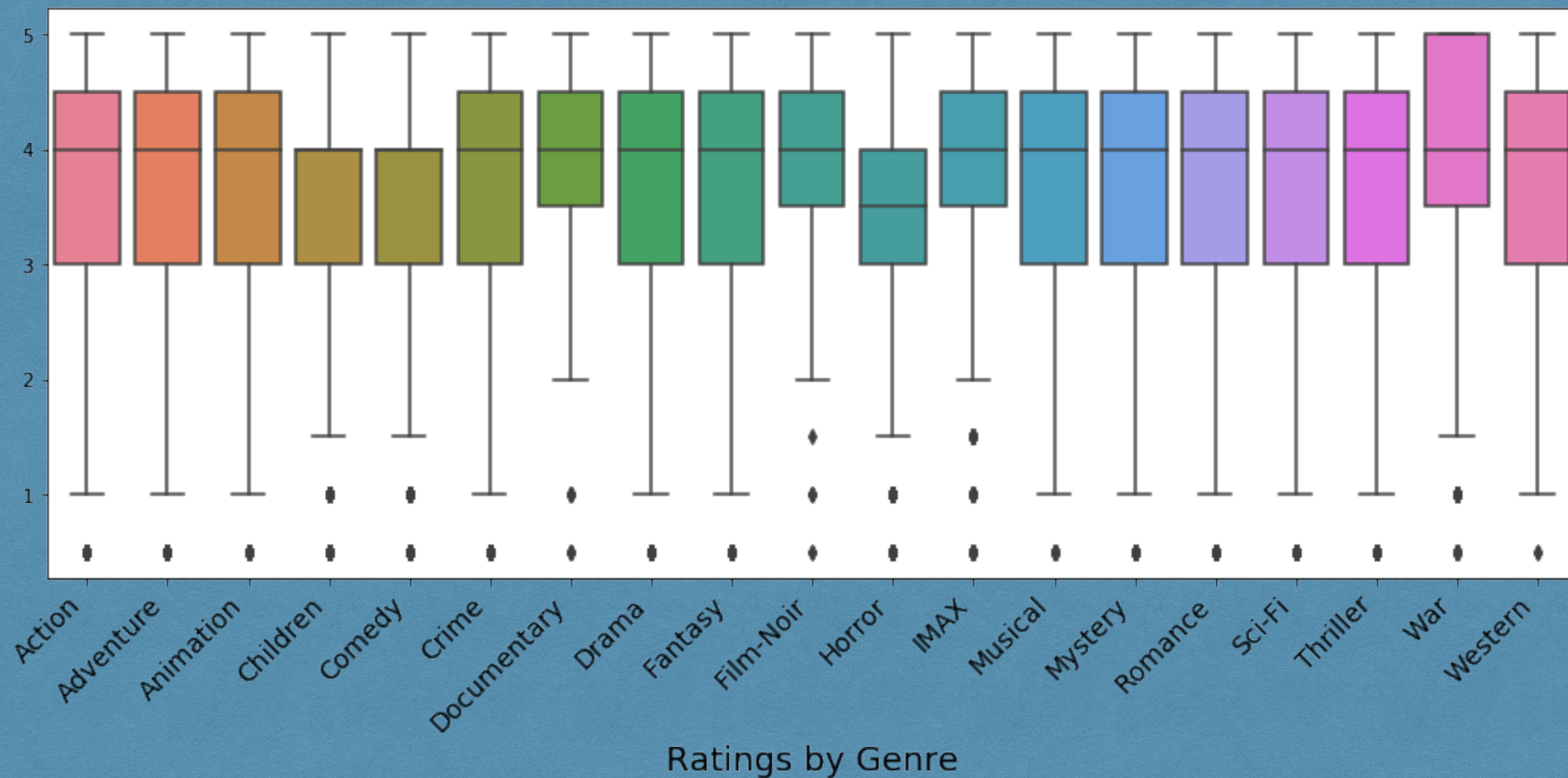

Movie Count per Genre (Can Overlap Genres)

The movies dataset was grouped by genre and the movie counts plotted

Visualizing the Data

Average Rating per Genre

The movies and ratings datasets were joined, grouped by genre and the average movie rating plotted
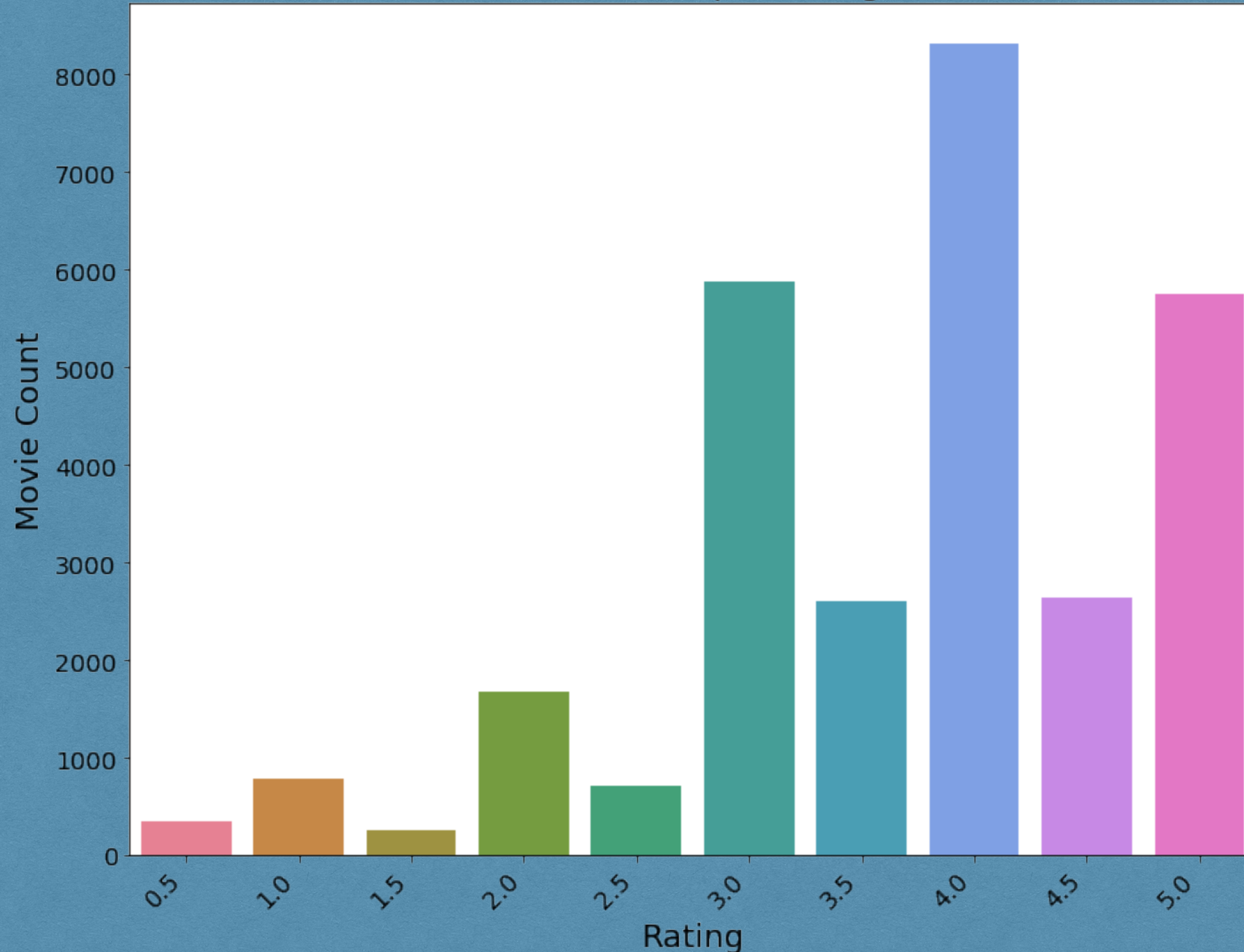
Visualizing the Data

The movies and ratings datasets were joined, grouped by genre and the movie ratings plotted
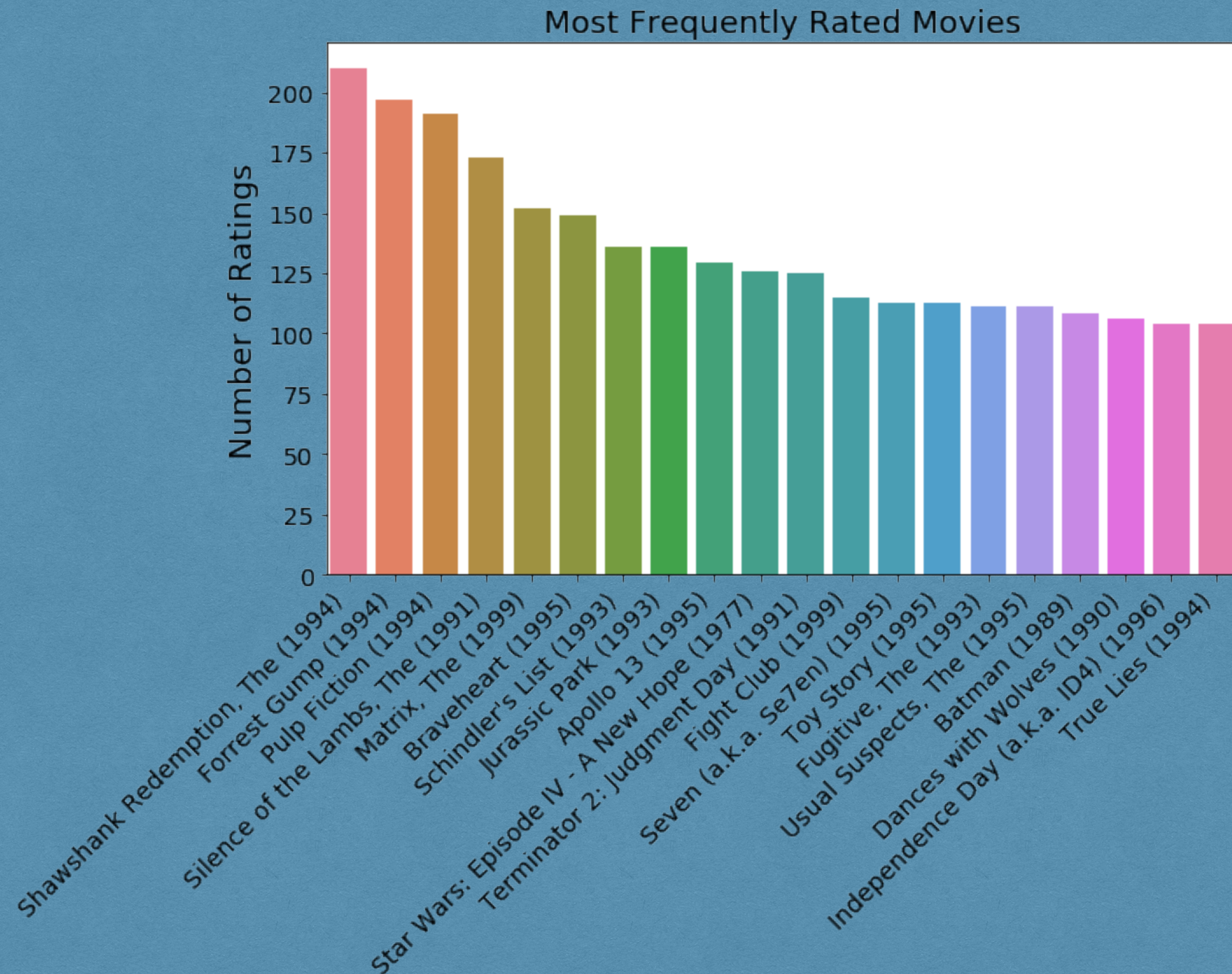
Visualizing the Data

Movie Count per Rating

The ratings dataset was grouped by rating and the movie counts plotted

Visualizing the Data

Most Frequently Rated Movies

The movies and ratings datasets were joined, grouped by title and the most frequently rated movies plotted

Building a Model

Recommendation systems were built using a python library called surprise

A number of algorithms within surprise were tested and scored

The top performing model predicts movie ratings for users with an average of ± 0.7 stars from their actual rating according to the mean absolute error

Avoiding the Cold Start Problem

New users are prompted to rate 5 movies from a list of the
500 most frequently rated movies

Making Predictions

The recommendation system then recommends 5 movies for the new user

Alterations to the Current Model

Could filter by genre, including a children's only dataset

Could filter by age of movies

Create dataset of current box-office movies

Thank You

Links for more information:

MovieLens:

https://grouplens.org/datasets/movielens/

Surprise Documentation:

https://surprise.readthedocs.io/en/stable/index.html