

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green color. Both are tilted at an angle.

Personalizing Movie Recommendations

Identifying Popular Genres and Tags to Expand Our Library



Overview

- Objective
- Data
- The Most and Least Common Genres in the MovieLens library
- Which Genres Dominate the Ratings?
- How Ratings Vary Across Different Genres
- Understanding the distribution of Ratings by Genre
- Improving Collection Diversity: Suggesting Genres
- Making the right recommendation for a recommendation model.
- Actionable recommendations
- How can we improve?
- Contact info



Objective

- **Create a movie recommendation system**
- **Check for completeness of FlixGenius library**
- **Recommend direction for adding new movies.**

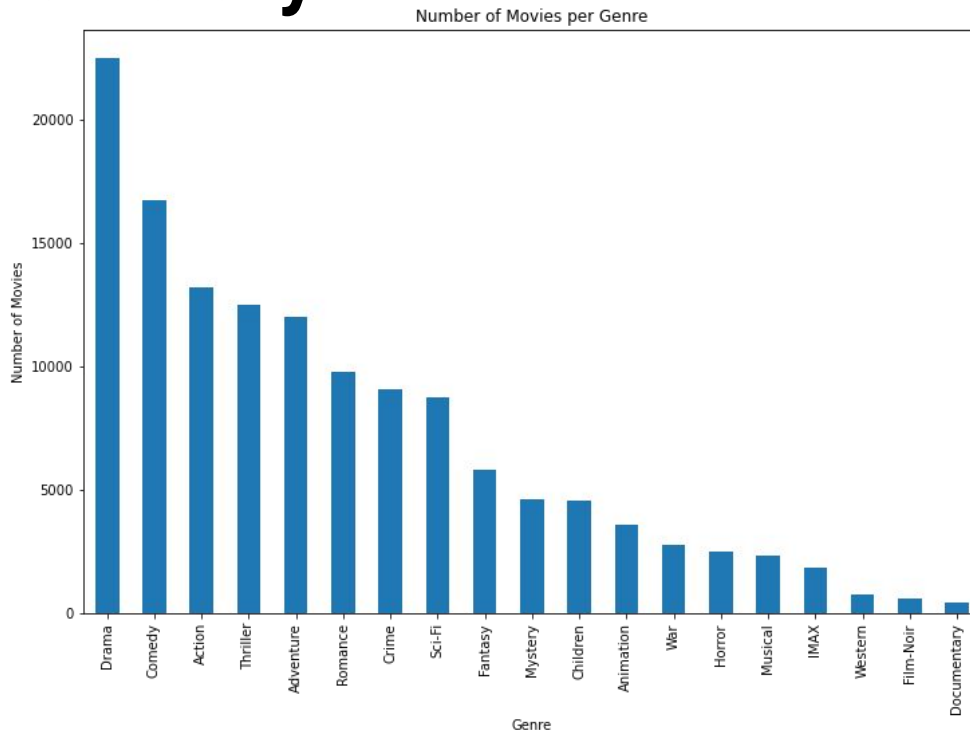


Data

- Dataset name: ml-latest-small
- Data description:
100836 ratings and 3683 tag applications
- 9742 movies
- Generated by 610 users between March 29, 1996, and September 24, 2018
- GroupLens Research: a research group in the Department of Computer Science and Engineering at the University of Minnesota that operates MovieLens, the source of these data.

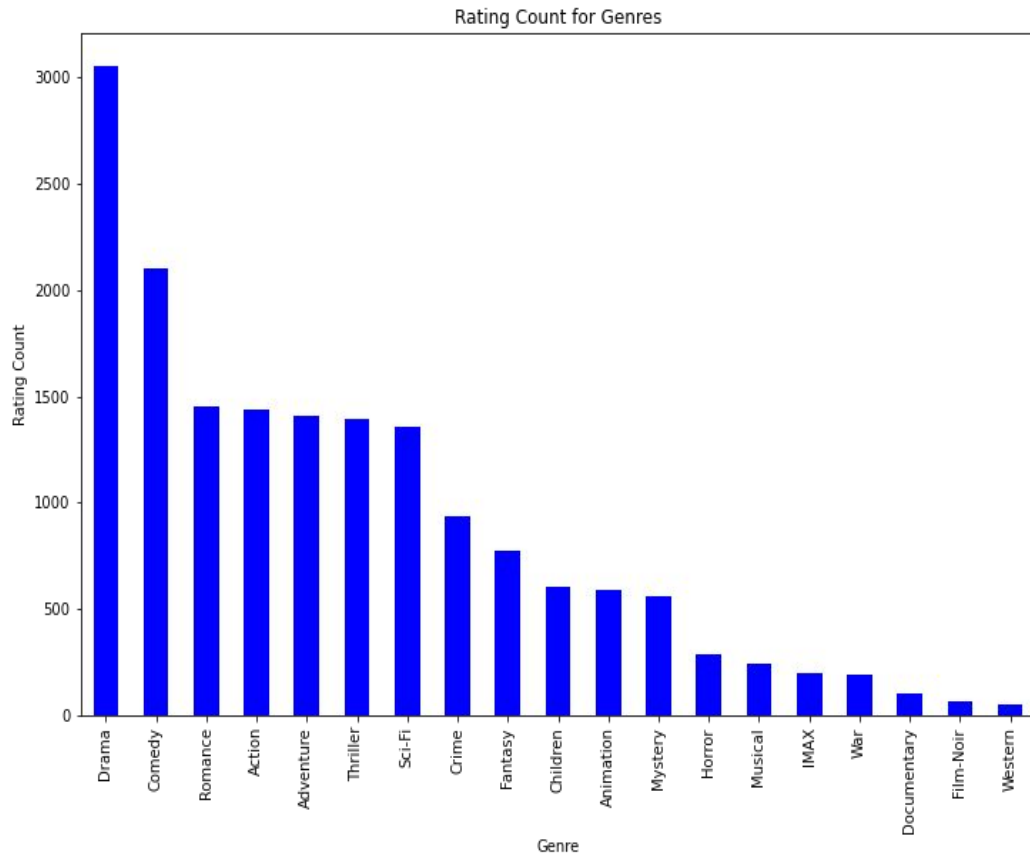
The Most and Least Common Genres in the MovieLens library

- Drama has the most number of movies
- Western, film-noir, and documentary are the 3 least represented genres



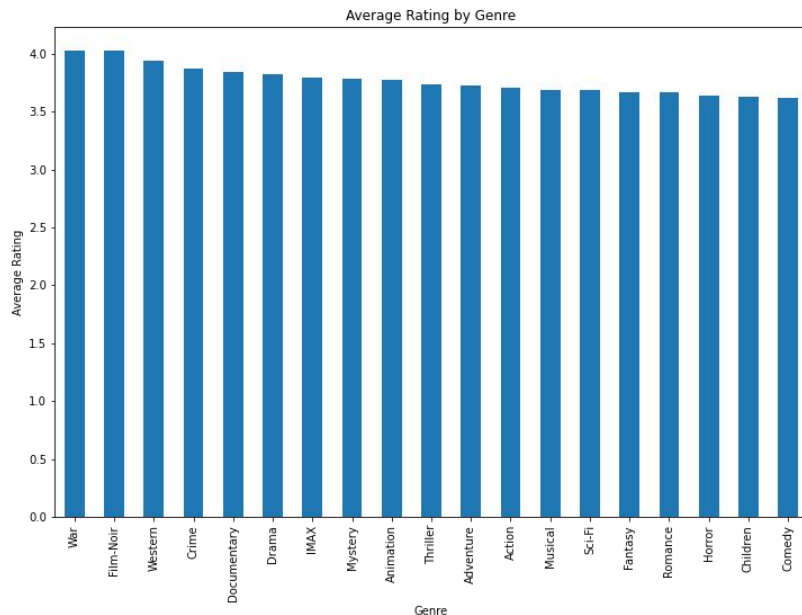
Which Genres Dominate the Ratings?

- Drama has the most ratings
- Documentary, Film-Noir, Western, and War are the least rated genres
- The least amount of ratings per genre resembles the number of movies per genre.



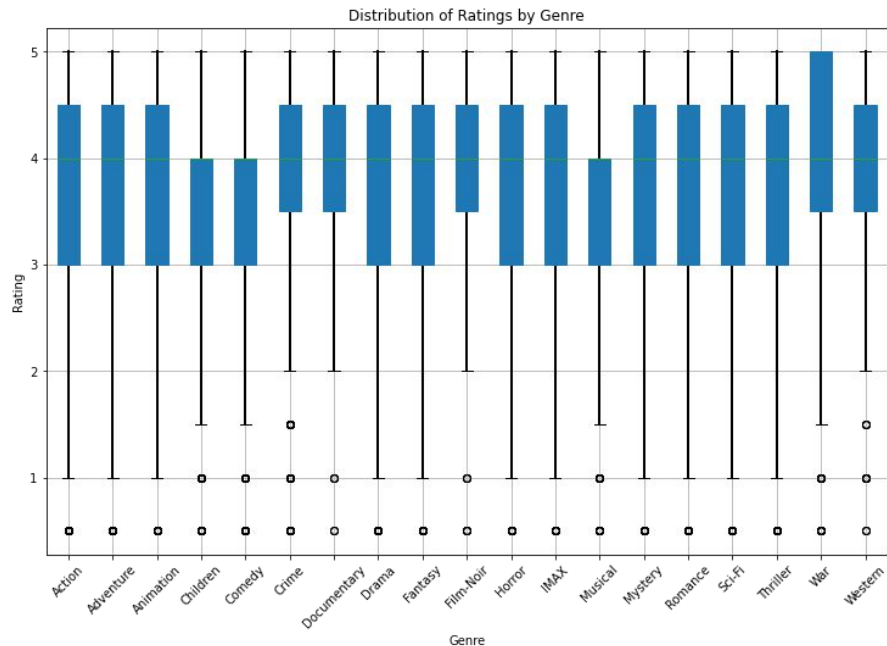
How Ratings Vary Across Different Genres

- Average ratings by genre are all between 3.5 and 4
- All genres are highly rated on average
- Any genre we expand on will be received well by our users.



Understanding the distribution of Ratings by Genre

- Our 4 focused genres all have a higher distribution of ratings.
- Documentaries, Film-Noir, Western, and War movies all have a higher distribution of ratings
- Fewer people watch them, but they are rated higher on average.
- War genre has the best distribution of ratings.





Improving Collection Diversity: Suggesting Genres

Based on the box and whisker plot of ratings, the war genre has the highest distribution of ratings, yet the least amount of frequency in our library according to the bar graphs. Therefore, it may be a good idea to add more movies from this genre to the collection. Additionally, genres such as western, film-noir, and documentary could also be explored further to enhance the diversity of the collection.



Making the right recommendation for a recommendation model.

- 3 types of recommendation systems were trained, collaborative being most accurate.
- Adding more data improved the accuracy of our model
- Root Mean Squared Error of 1.18 indicates that, on average, the movie recommendation system is making predictions that are about 1.18 rating points off from the user's actual ratings
- An MAE of 0.36 indicates that, on average, the movie recommendation system is making predictions that are about 0.36 rating points off from the user's actual ratings

Baseline Collaborative Model

Mean RMSE: 1.9174329156295056
Mean MAE: 1.404918032786885

Updated Collaborative Model

RMSE: 1.1983601518382785
MAE: 0.40154738878143137

L2 Ridge Regression Collaborative Model

RMSE: 1.180387442535488
MAE: 0.36871145409131517



Actionable recommendations

- Deploy our new Collaborative Based Filter.
- Add more movies, specifically from the lesser populated genres.
- Add more data. More data such writers, actors/actresses
- Create a plan to encourage users to rate more movies.
- Stress to users that the more movies they rate and watch, increases the odds of getting recommended movies they'll love



How can we improve?

- More data!
- Collect more personable data, something like age can seriously impact what someone wants (or is allowed) to watch.
- Look into revisiting the rating system, A larger scale (1-10) could possibly increase the accuracy of our model.



Thank you for your time.

We hope this presentation has given you confidence in our Collaborative model.

We are excited to continue striving to make our model more accurate, and to recommend movies to people that they will love!



Feel free to contact me at
MichalOst3389@gmail.com