Module 4 - Final Project

Will Dougherty
Flatiron School - Data Science - Online

A Brand New **FlixWeb**

A new approach to
User Recommendations

Help our users answer:

“What do you want to watch?”

A Dual Approach

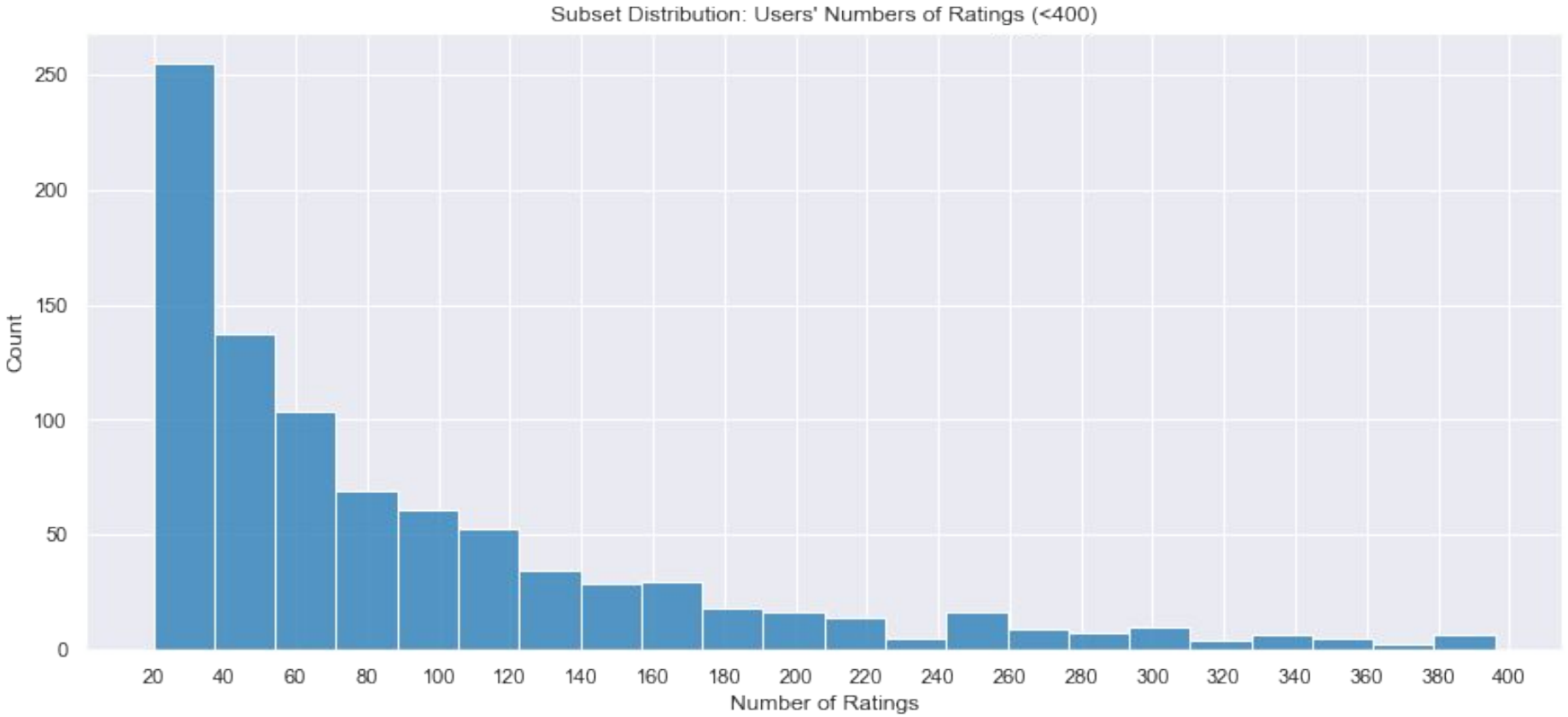
New Users

- 0-20 Ratings
- Top Films at first
- Content-based filtering
- Recommend “Movies Like This” based on past views and ratings
- Focused on discovery and building user profile

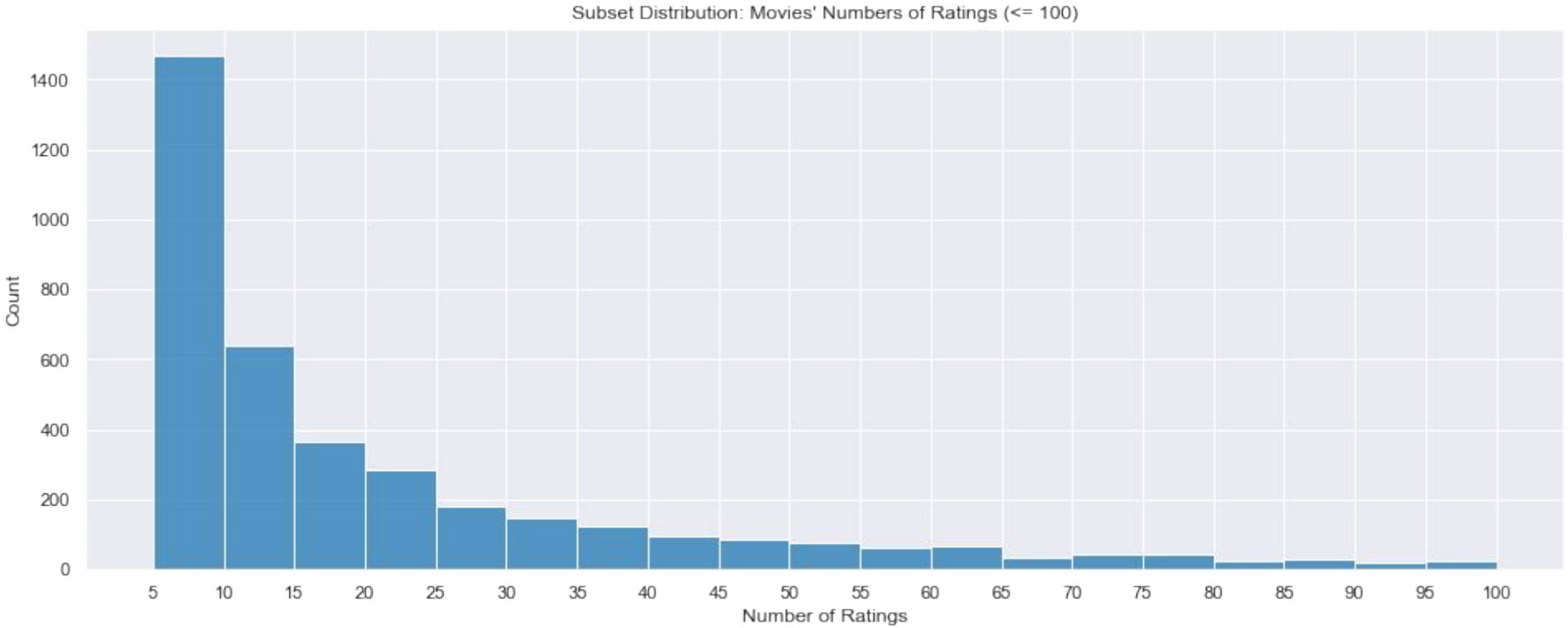
Mature Users

- 20+ Ratings
- Collaborative Filtering
- Highly Personalized
- Focused on recommending films that similar users enjoyed

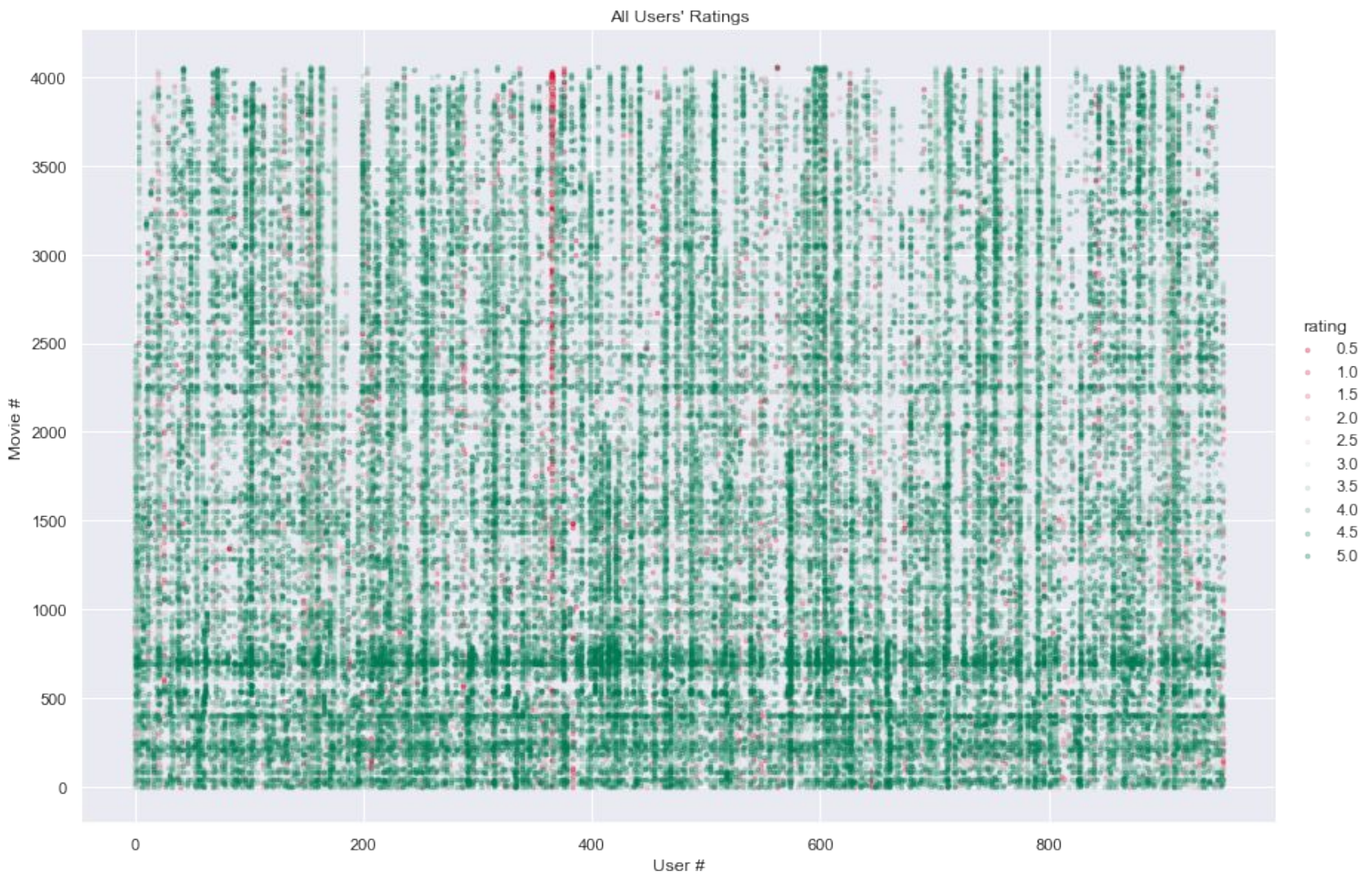
Dataset and Methodology



900 Users, 20+ Ratings each



4000 Movies, 5+ Ratings Each


All Users' Ratings
Movie #
User #
0
500
1000
1500
2000
2500
3000
3500
4000
0
200
400
600
800
rating
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0

New Users

New users are provided top films:

- All films
- By Genre
- By Decade
- Hidden Gems (< 15 ratings)

After one rating:

- Content-based Filtering
- Finds films similar to users’ recent views
- Uses genres, decade, and user-generated tags

‘Mature’ User Profiles

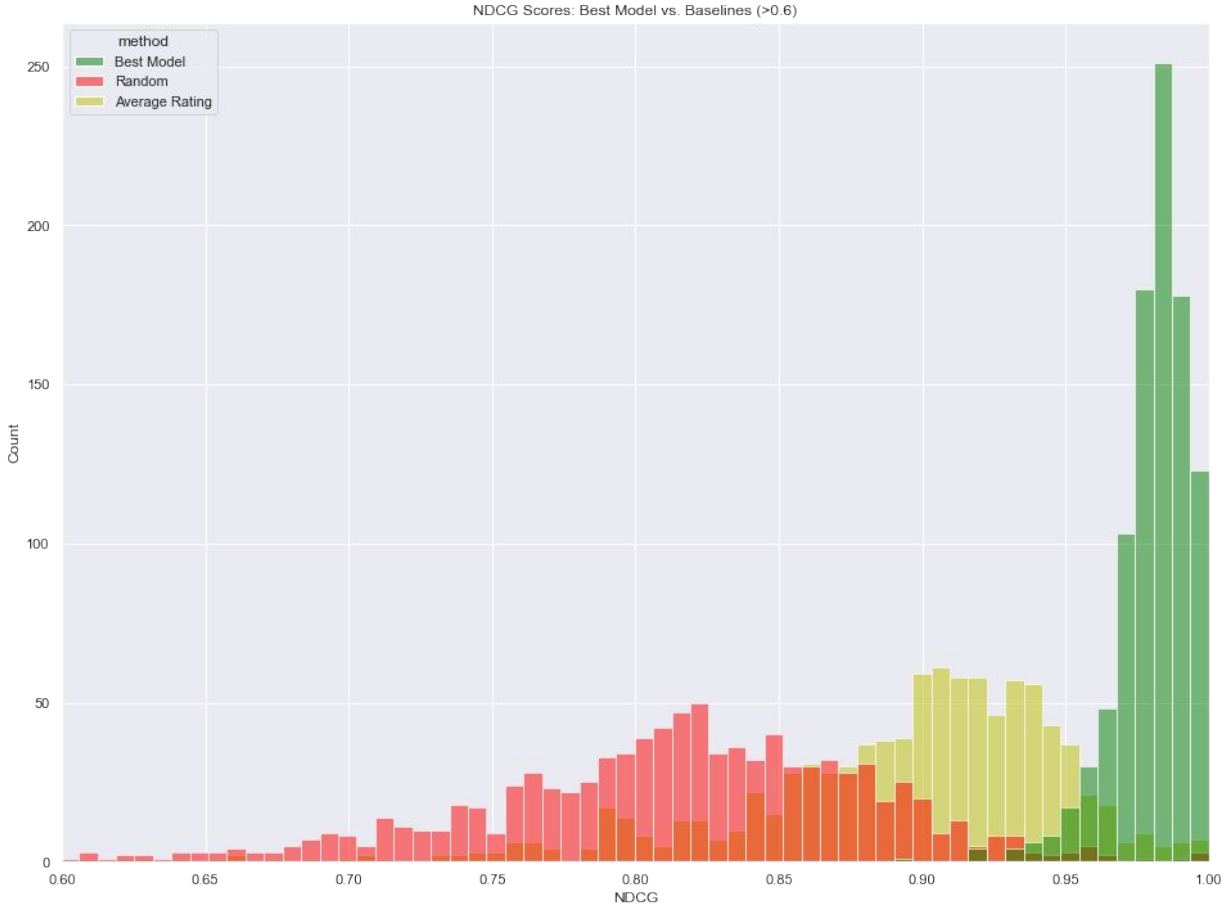
After 20 ratings, the user’s profile is folded into the main model, and predictions are generated.

- ALS Collaborative Filtering Model
- Predicts user’s rating of a given movie based on similarity to other user profiles
- This is used to rank user’s top-rated films

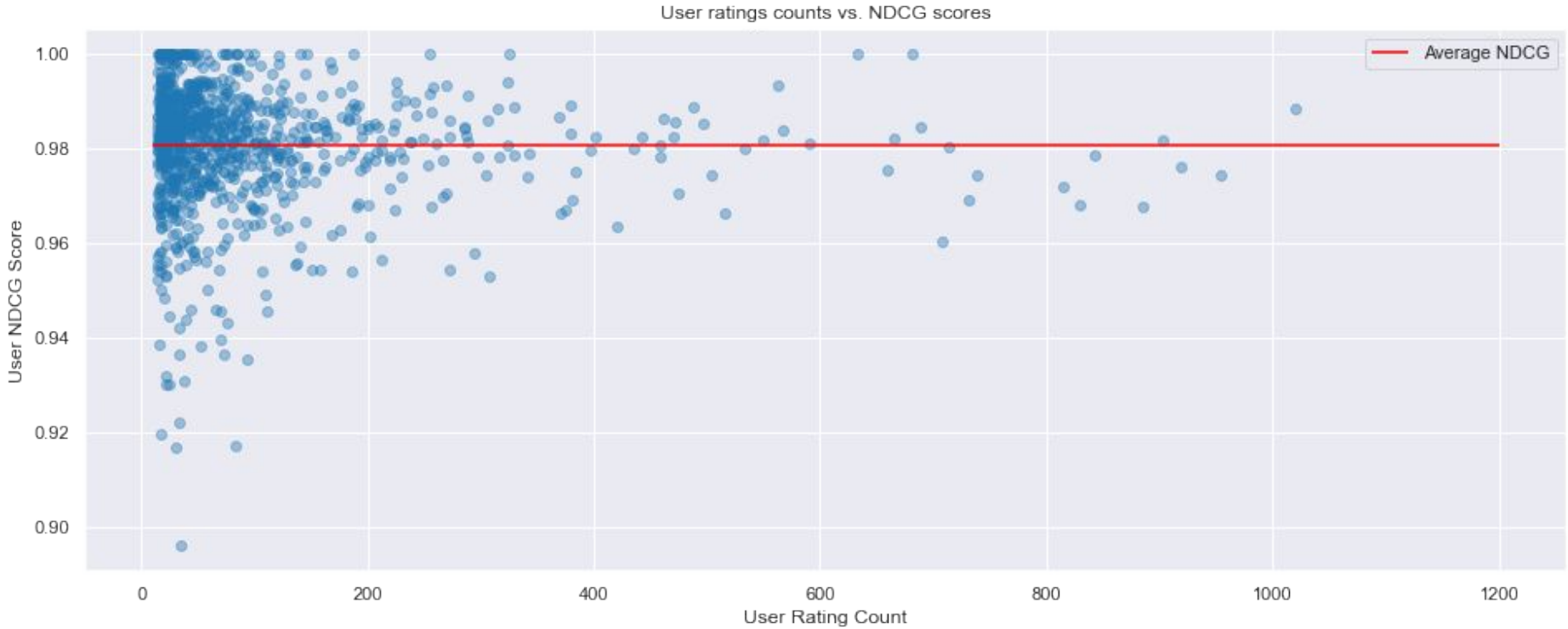
This can generate:

- Top 5

- Subsets by:
  - Genre
  - Decade
  - Hidden Gems



Evaluation of Collaborative-filtering Model
NDCG Scores - How Well the Model Ranks Films



User NDCG Scores vs. Rating Counts

Benefits of This Approach

- **Applicable to All Users**
  - New users
  - Current users
- **Better recommendations = More Engagement**
  - Less time searching for a movie to watch
  - Users will have a better experience, and come back for more content that speaks to them

Future Work

Refine content-based system

- Recommend based on multiple films
- Refine / Develop tags and metadata for content-based system

Develop new features for both systems

- i.e. recommend cross-sections of genre / decade if there is a strong pattern in a user profile

Thank you!