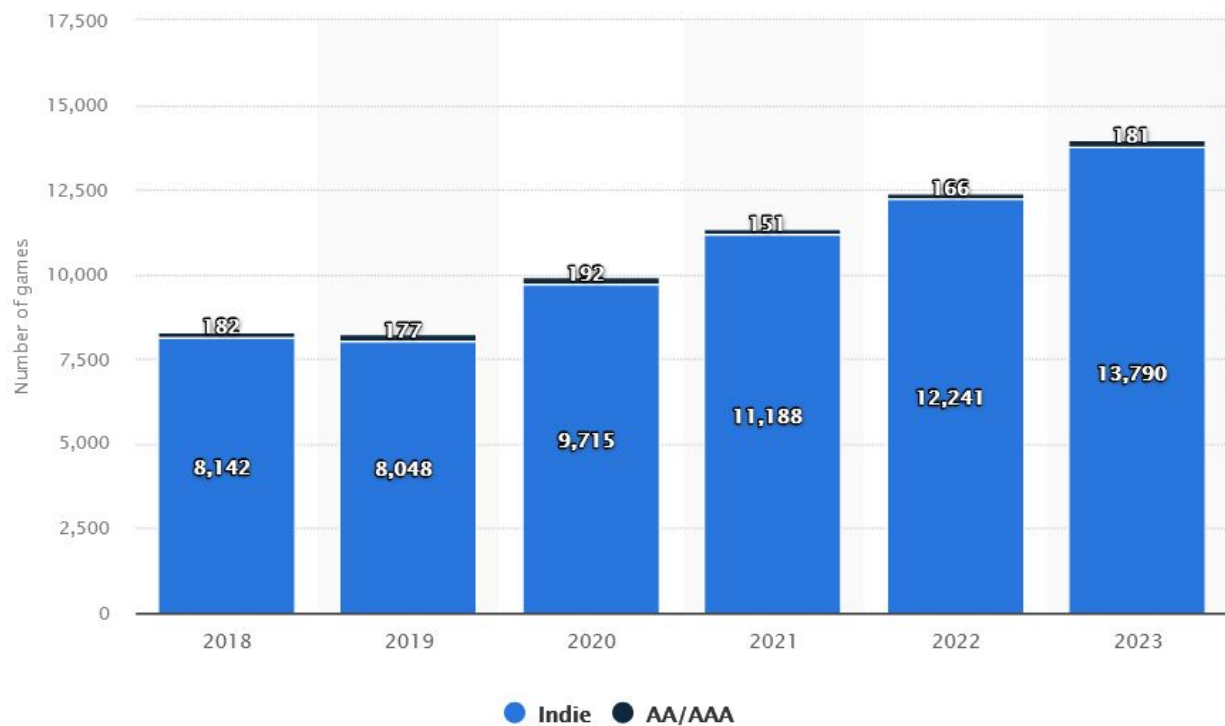# Indie Game Recommendation System

Zachary Cohen's Capstone Project
Sprint 2

Presentation Overview

1. Intro
2. Approach
3. Exploratory Analysis
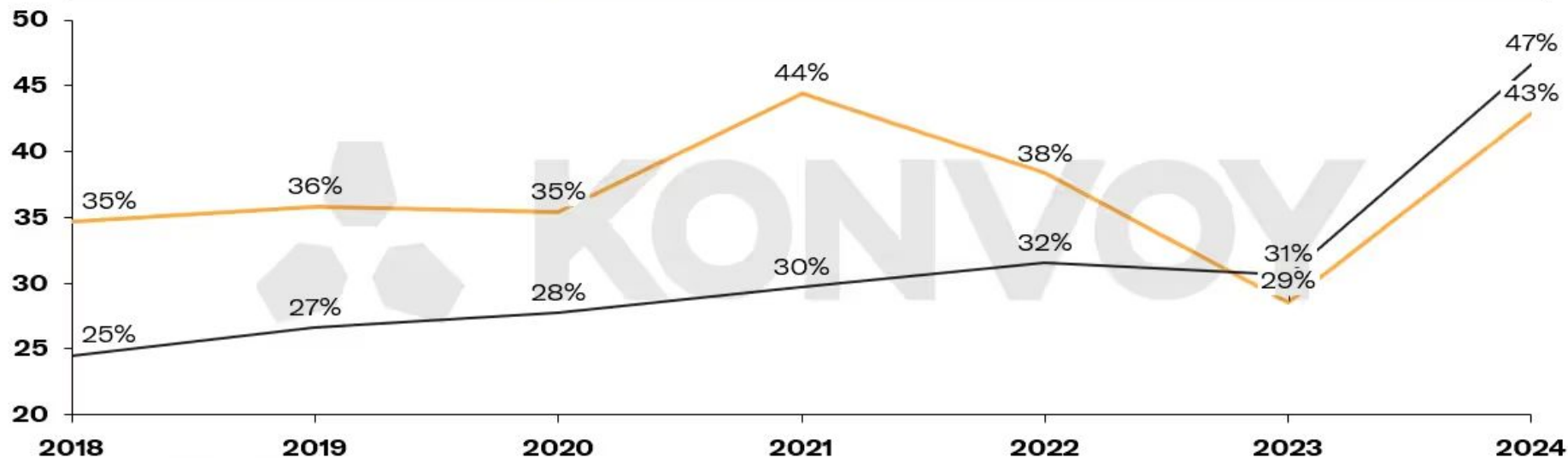4. Modeling
5. Results

Introduction

https://www.statista.com/statistics/1411839/number-games-released-steam-developer-type/

Indie game revenue on Steam — Indie Games as % of Games >$10M Rev — Indie Games as % Steam Rev

Year	Indie Games as % of Games >$10M Rev	Indie Games as % Steam Rev
2018	35%	25%
2019	36%	27%
2020	35%	28%
2021	44%	30%
2022	38%	32%
2023	29%	31%
2024	43%	47%

Source: VG Insights

https://www.konvoy.vc/content/the-era-of-the-indie-game

Impact approximation

132 M active users/month * 1% * $20 =

$264,000/month per percent of monthly users who purchase a game
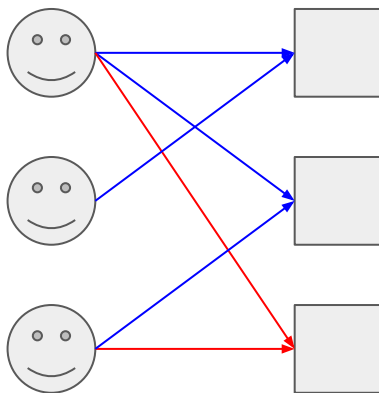
$79,000/month (30%) to Steam

$184,800/month (70%) to indie developers

(Monthly active users from 2021 data by Statista)

Approach

Recommender Approach

- Find latent game/user categories
  - Derived from review graph
  - Derived from topic analysis of reviews
- Recommend most popular unknown games in game category

~~Recommender Approach~~ Classifier Approach

- Find latent game/user categories
  - Derived from review graph
  - ~~Derived from topic analysis of reviews~~
- ~~Recommend most popular unknown games in game category~~
- Predict review sentiment based on user/game clusters
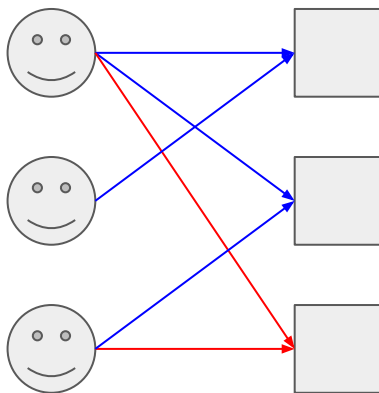
Classifier Steps

1. Identify communities
   - Train/Test Split -
2. Pre-process
3. Calculate sentiment probabilities per community combination
4. Logistic Regression

Classifier Steps

1. Identify communities
   - **Train/Test Split -**
2. Pre-process
3. Calculate se   s per com
4. Logisti

Exploratory Analysis

Data

Review JSON Contents

- **Sentiment**
- Review text
- User information
  - Id
  - Game playtime
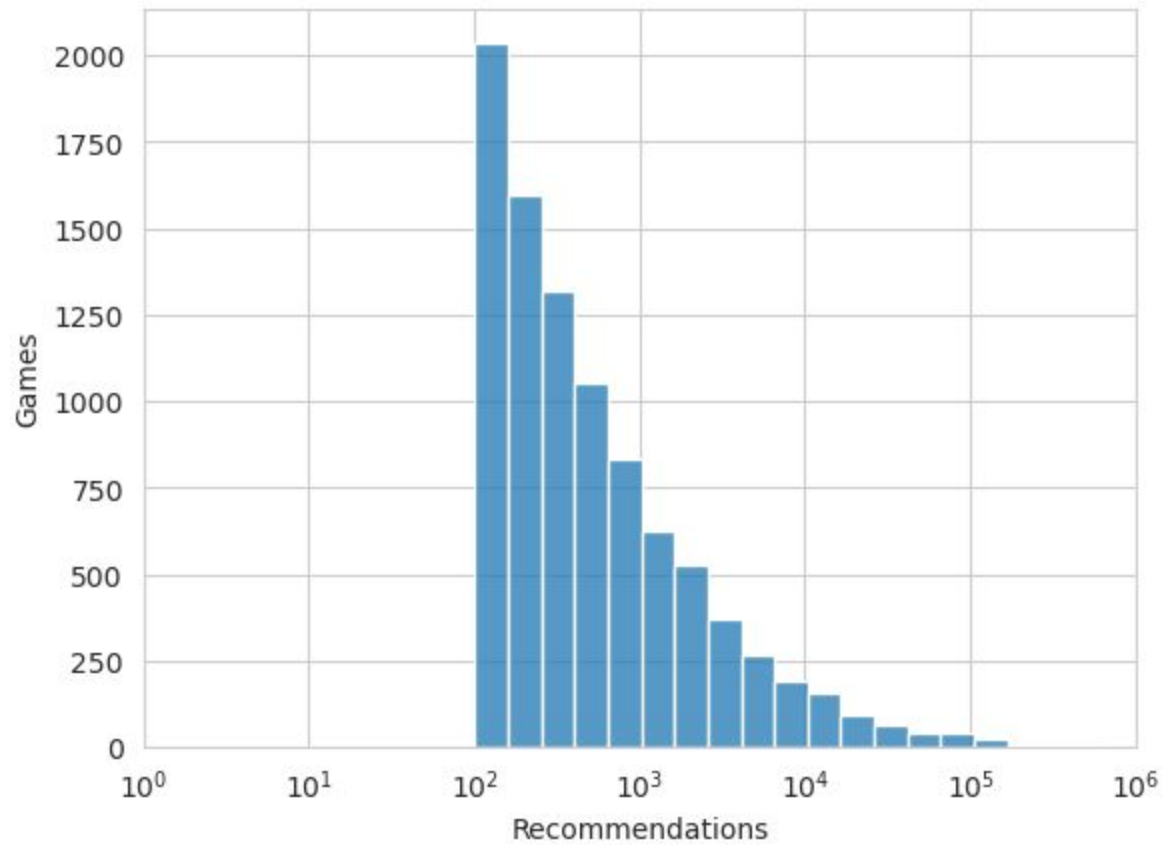   - At review
    - Total
    - Past two weeks
- Review votes

Game JSON Contents

- **Price**
- **Recommendations**
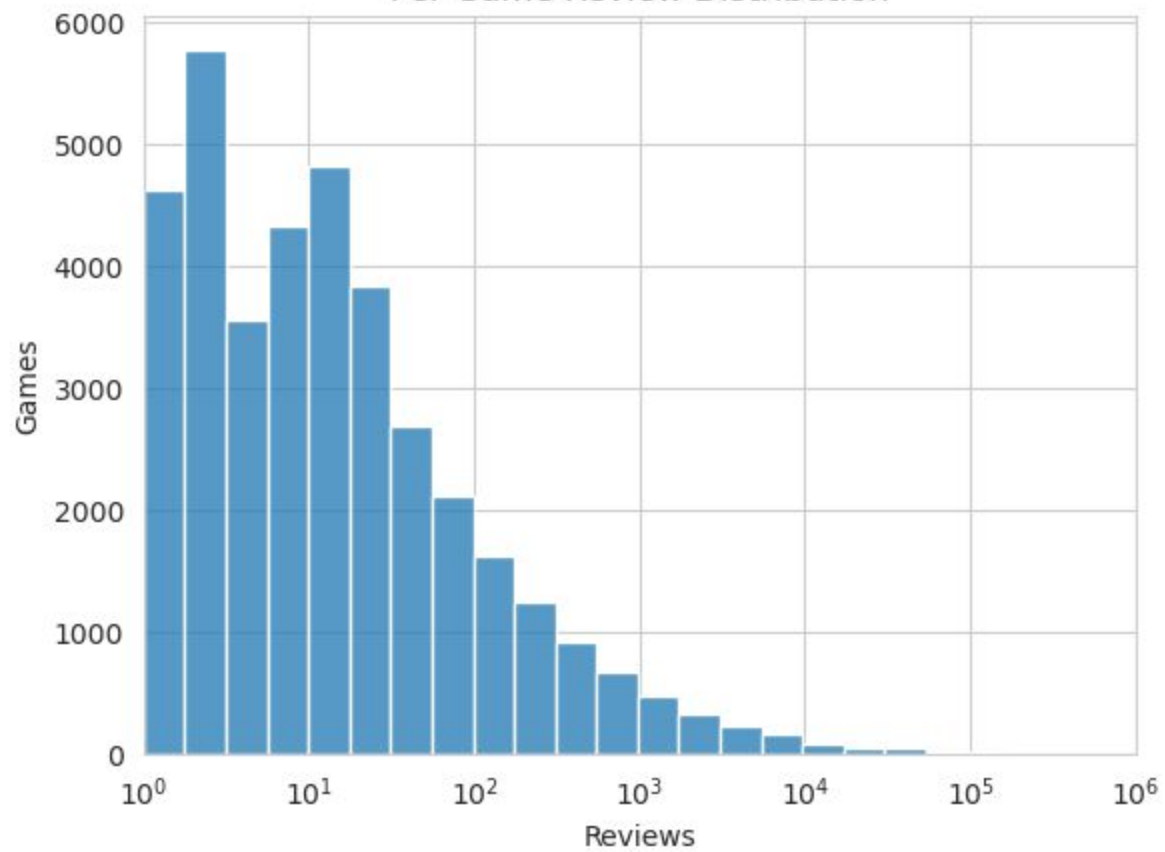- Genres
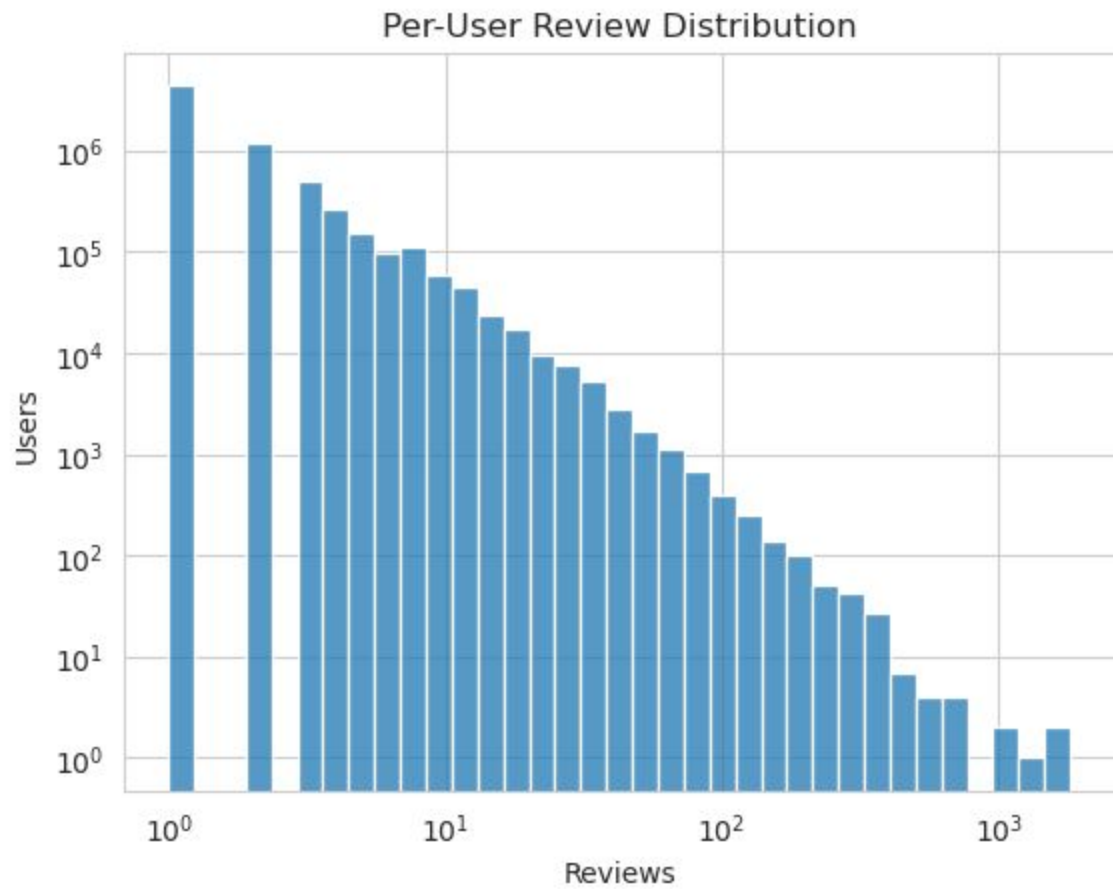- Compatibility
- Age

Scale
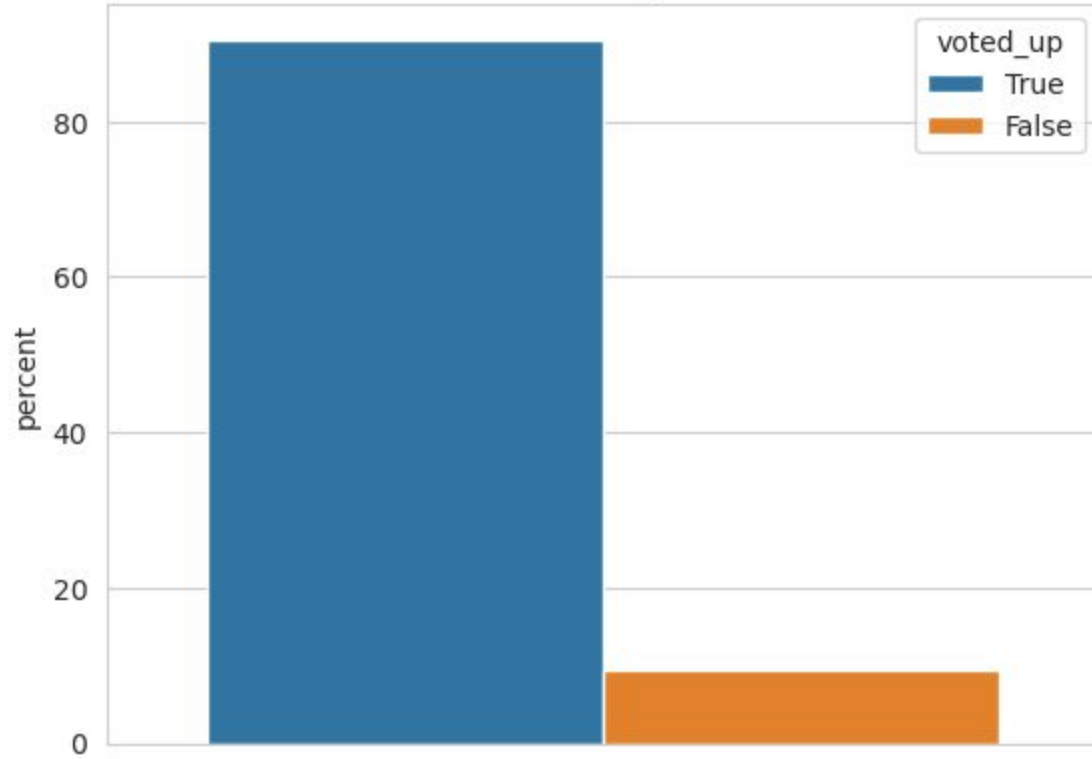
- 37,588 games
- 6,908,265 users
- 14,431,826 reviews

Per-Game Recommendation Distribution

Per-Game Review Distribution

Per-User Review Distribution
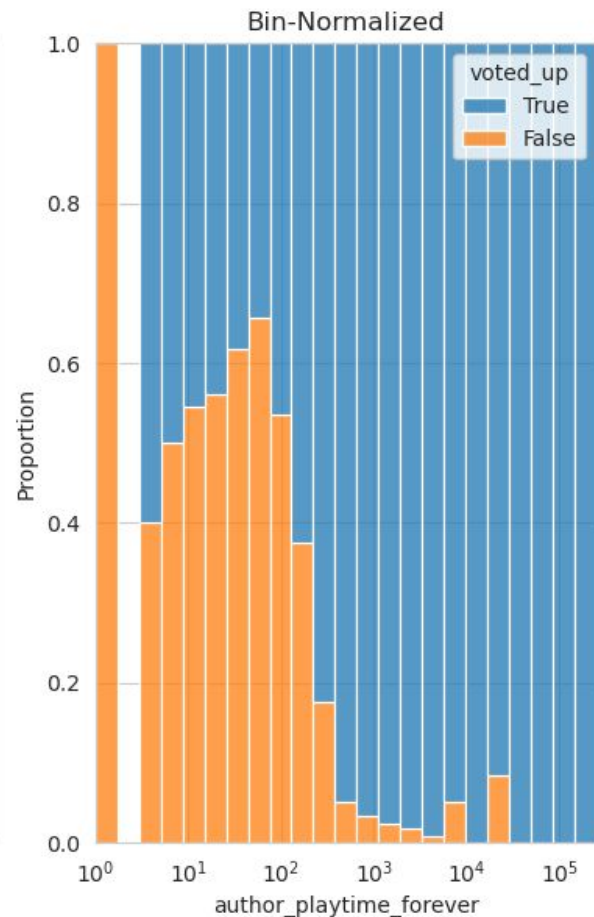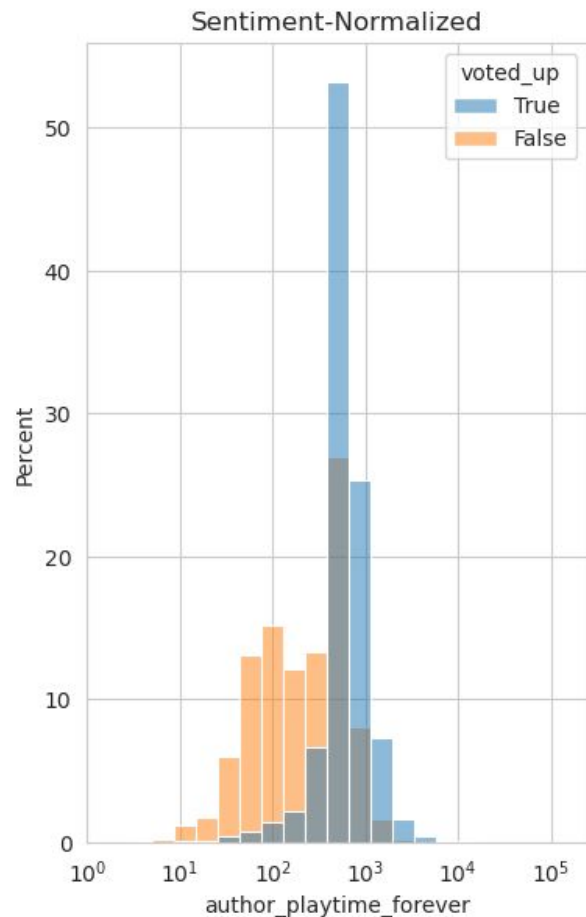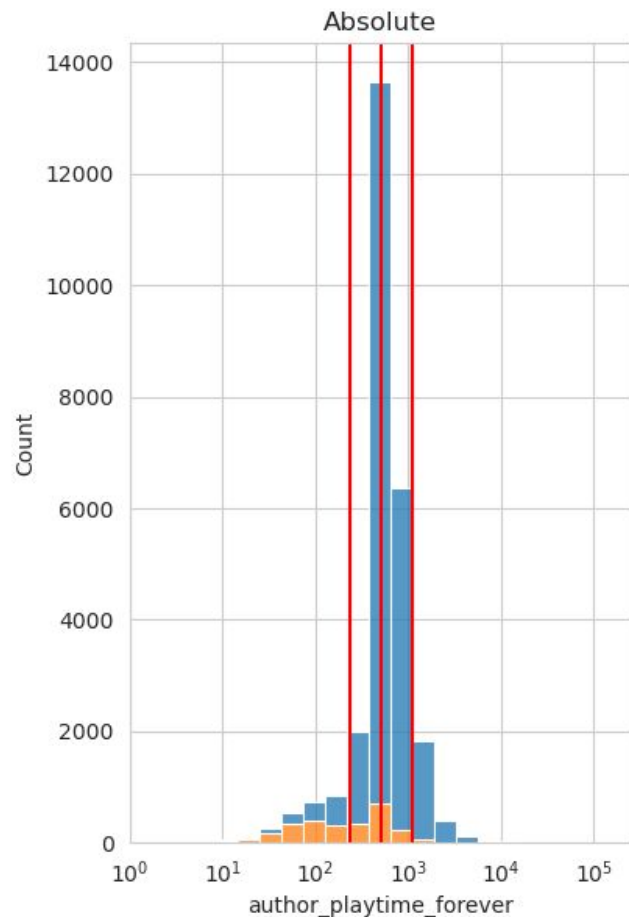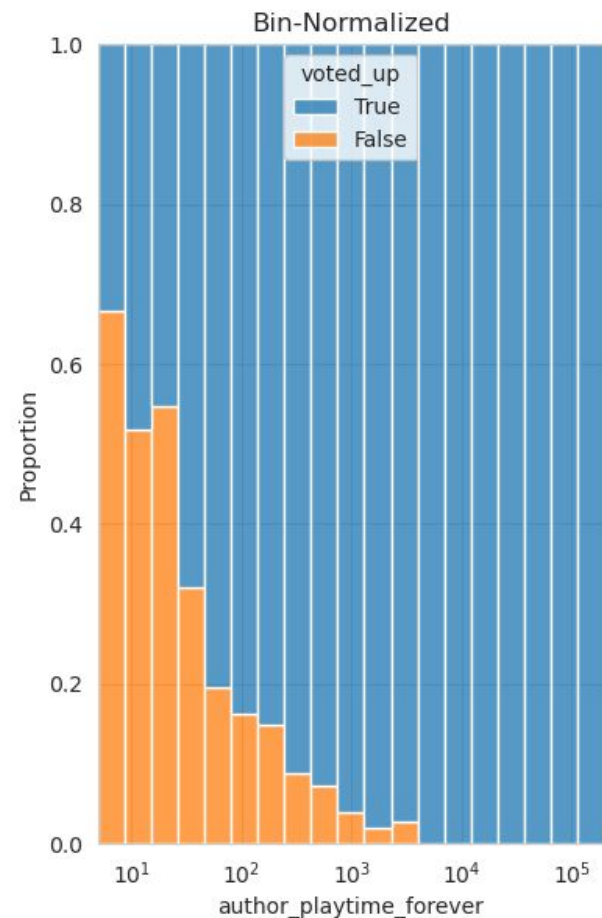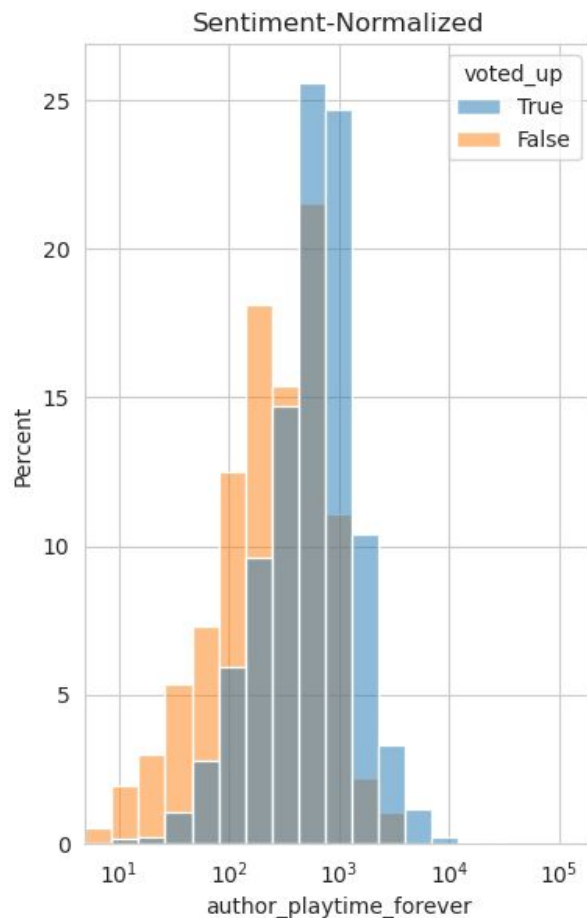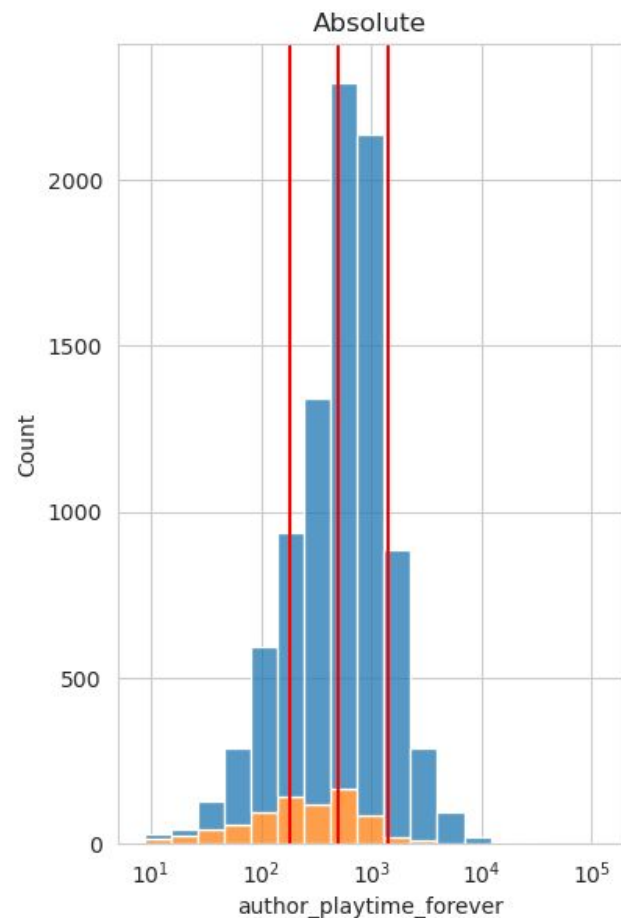
Review Percent by Sentiment

Hellblade: Senua's Sacrifice: Lifetime Playtime

Warhammer 40,000: Boltgun: Lifetime Playtime

Modeling

Community Clustering

- Two bipartite graphs: positive and negative reviews
- Communities identified via Leiden algorithm (scikit-network)
  - Communities with < 5 members merged together
  - Users/games with no community all assigned to the same community

	Percent Total		Absolute		
	Positive	Negative	Positive	Negative	Total
Games	94	74	35,181	27,971	37,588
Users	95	13	6,554,623	883,054	6,908,265
Reviews	91	9	13,074,002	1,357,824	14,431,826

Cluster Summary

	Count		Median Member Count	
	Positive	Negative	Positive	Negative
Games	**48**	**1066**	109	14
Users	**126**	**1258**	10	419

Cluster Summary

	Count		Median Member Count	
	Positive	Negative	Positive	Negative
Games	48	1066	**109**	**14**
Users	126	1258	**10**	**419**

Pre-processing

- Price, recommendation counts
  - Log-transformed; robust-scaled
- Previous review sentiment
  - Calculated as positive/negative review ratio; robust-scaled

Label probabilities

- Positive sentiment probabilities calculated per training set user label/game label combination
    - Handled separately for labels derived from positive/negative sentiment graphs
- Test set probabilities for unseen label combinations imputed as positive/negative probability midpoint
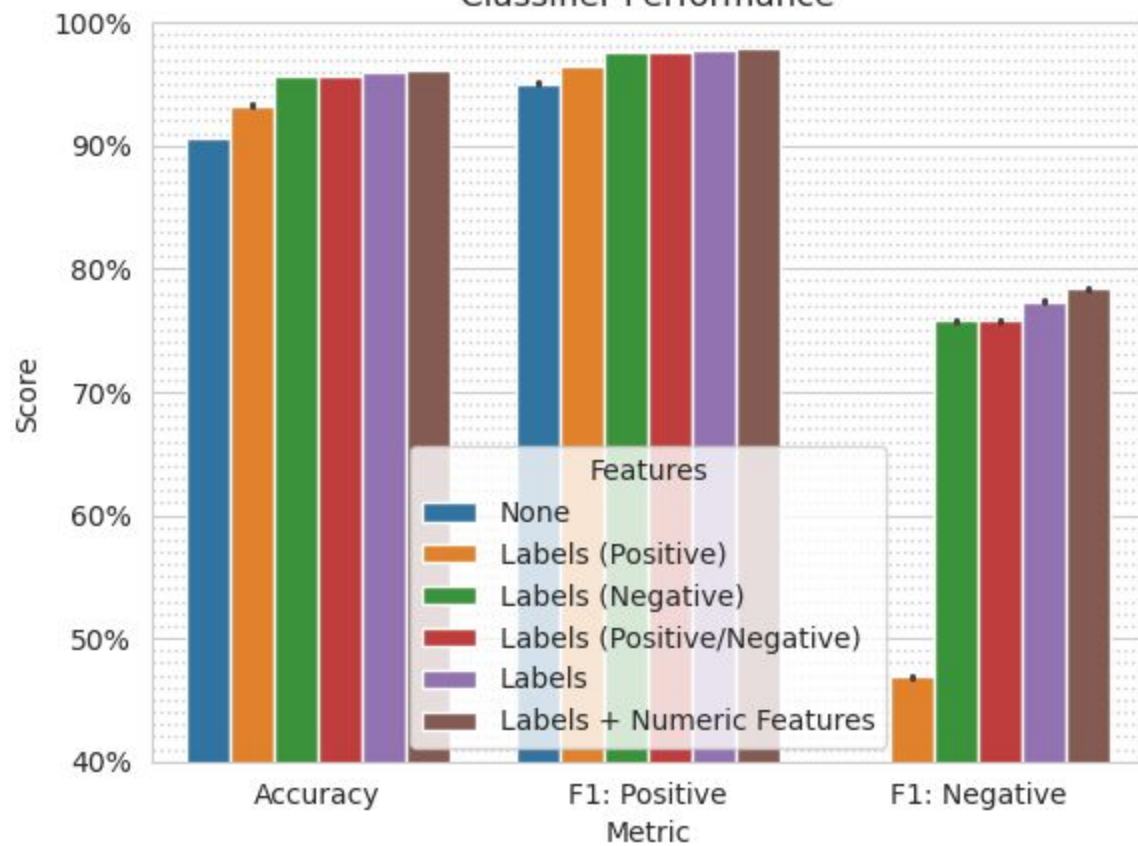   - (mean of positive and negative sentiment means)

Logistic Regression

- 5-fold cross-validation, no hyperparameter tuning

Results

Classifier Performance
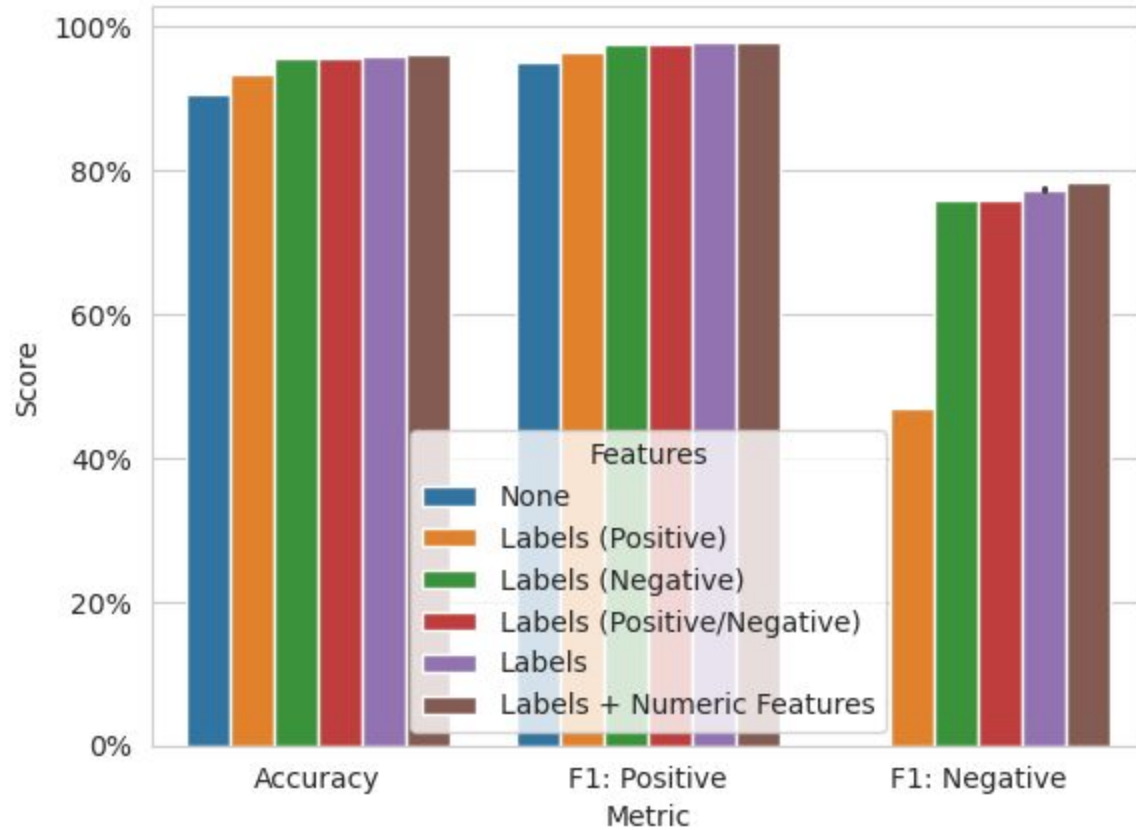
Classifier Performance

Next steps

- STOP PLAYING WITH GRAPHS
- Attempt to derive human-interpretable labels from review texts ("souls-like", "metroidvania", etc.)
  - Topic analysis
- Transition from classifier to recommendation system

Thank You