

Recommender System

Based on Steam dataset

Crew member : Alex, Sen, Cynthia , Lydia, Henry, Ben



The Team



Sen Zhang

Leader



Henry Yao

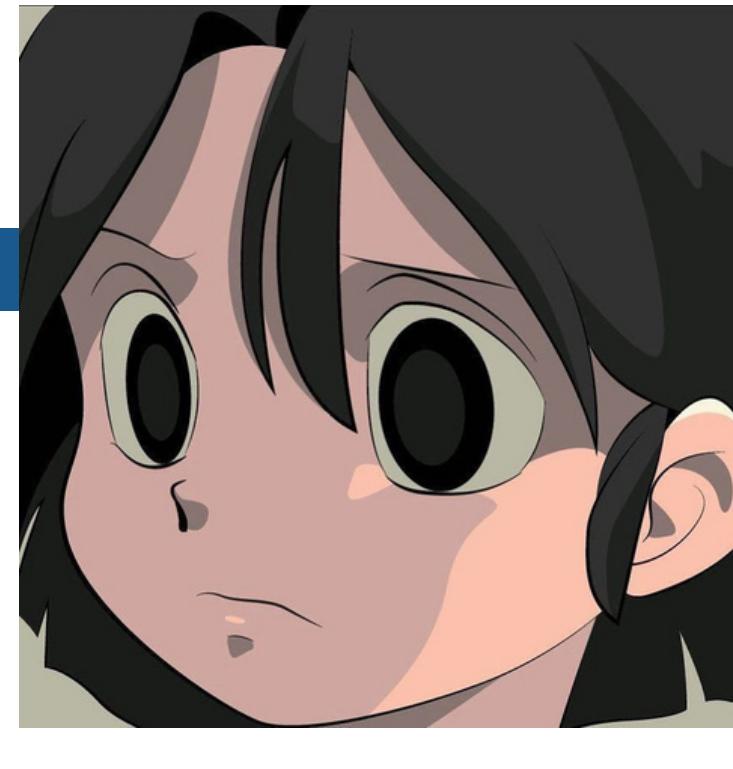
Editor



Alex Yuan

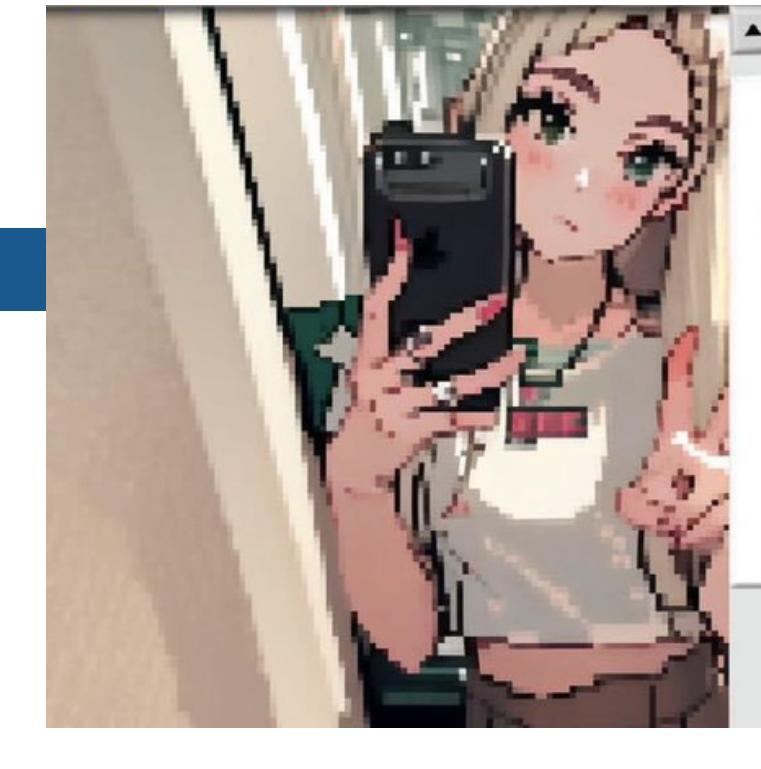
Programmer

The Team



Cynthia Tian

Researcher



Lydia Li

Researcher



BenLiu

Editor

Overview

- Introduction
- Problems
- Methods
- Model presenting
- Results
- Discussion
- Conclusion





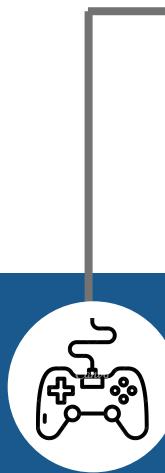
Introduction

Steam is a large digital distribution platform developed by Valve Corporation, a well-known video game company. It serves as a online store platform where users are able to purchase, download, and play video games on their PCs.





Steam Service



DIGITAL
DISTRIBUTION

WORKSHOP AND
USER-GENERATED
CONTENT



GAME STORE

SALES AND
DISCOUNTS



MULTIPLAYER AND
ONLINE SYSTEM

COMMUNITY AND
REVIEWS

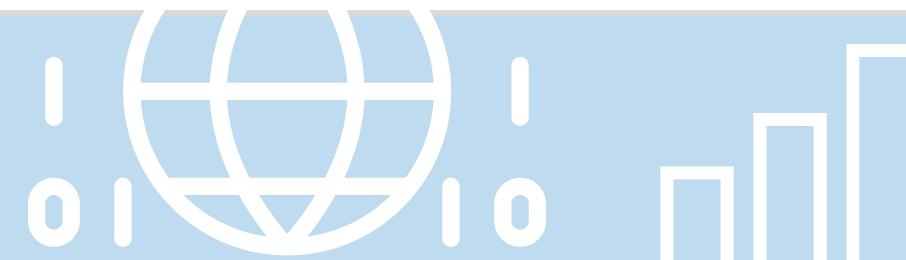


ACHIEVEMENTS AND
TRADING CARDS

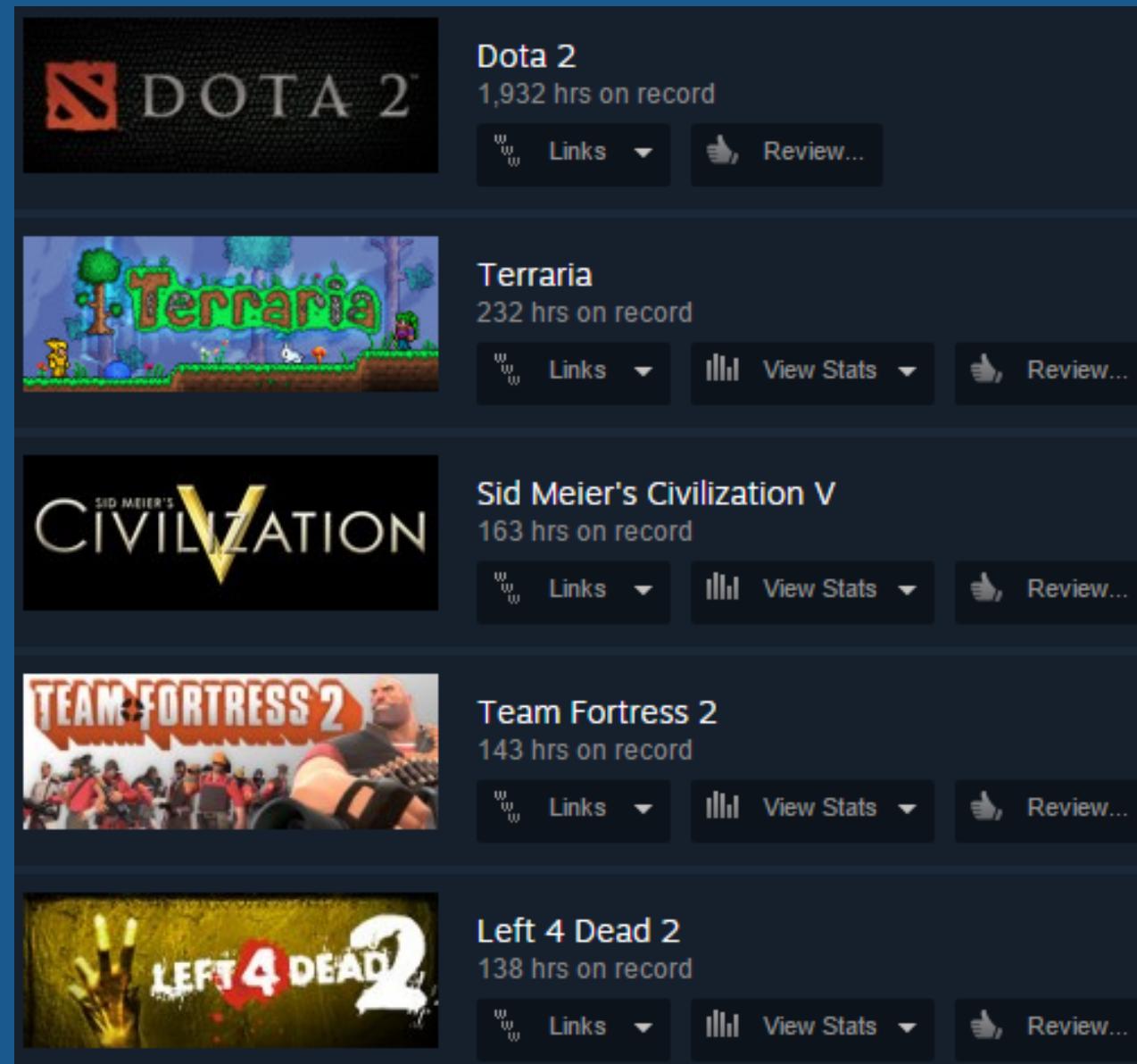
CROSS-PLATFORM
SUPPORT

Introduction

We utilized the data sets from Kaggle.com in order to discover user interactions with items(games). The datasets include, for example, the playing hours for each game of individual users, purchasing behaviors, game information, etc.



Introduction



We explored the performance of three distinct methodologies of the recommender system for Steam games. Algorithms based on each model are designed and developed in Python, with the use of Pandas Library.

Problems

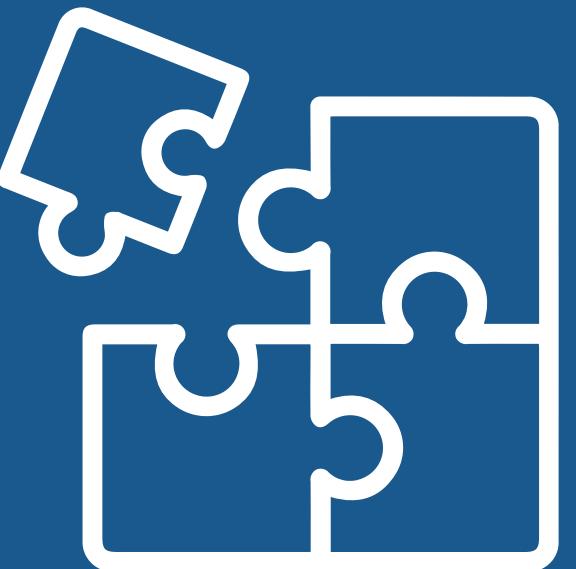


Current Steam recommender system:

- The “Popular Bias” phenomenon-Favoring those who are favored...
- ...while games from small companies not gaining enough credit
- Reviews written by the individual users does not fit all other users

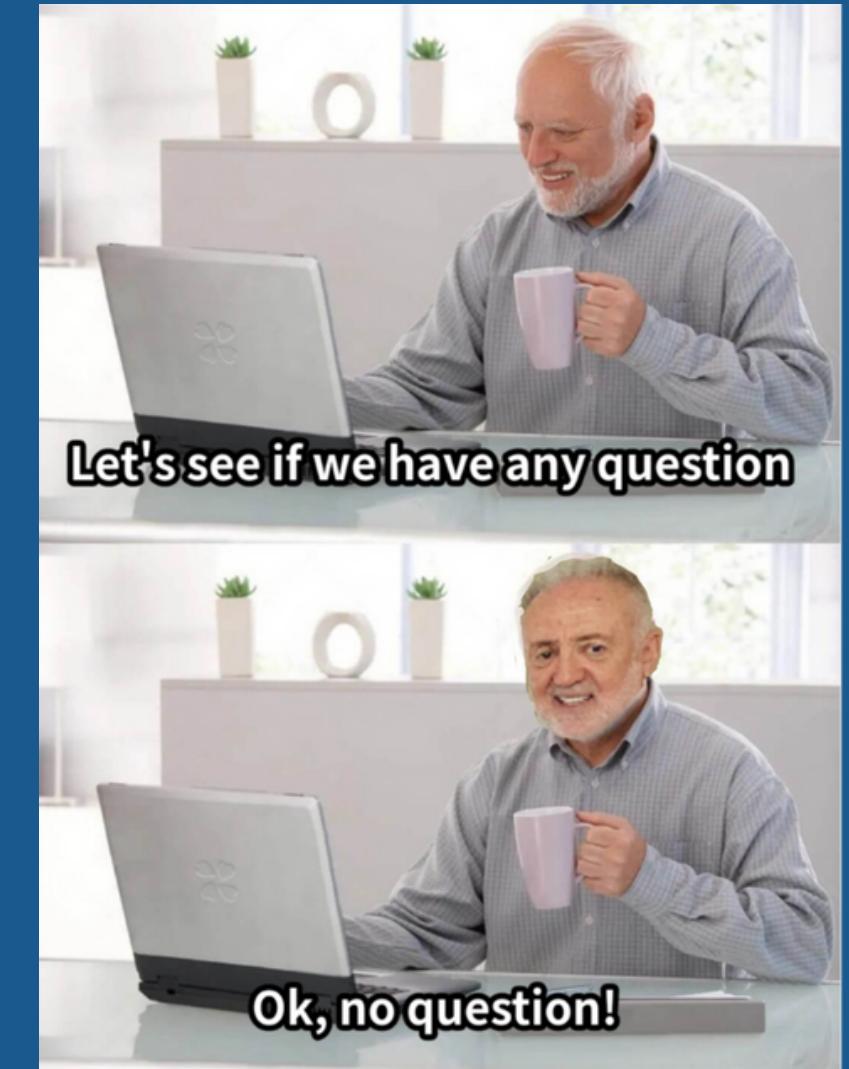
Solution

To gain a more accurate prediction of user preferences, our idea is to focus on user interaction behaviors that are directly related to their interests("purchasing history", "playing hours"). People would tend to spend more time on games that they are interested in, which allows playing hours a natural indication of preferences. Therefore, our systems are based on the playing hours and purchasing behaviors of users.



Solution

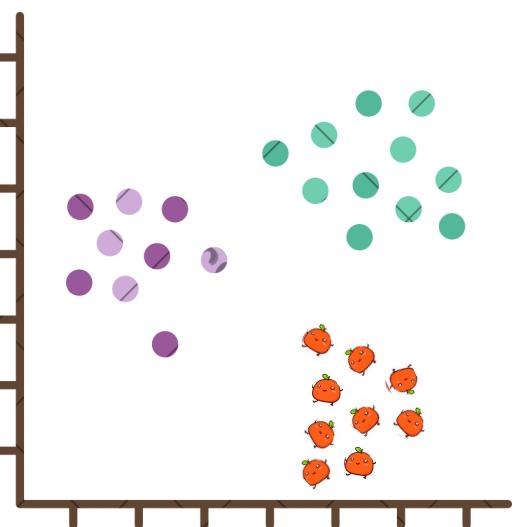
Also, different statistical models are utilized to gain more comprehensive knowledge in terms of improving the recommender system. We used K-means model, Gaussian Mixture Model(GMM) and item-based collaborative filtering respectively to build up algorithms. The three algorithms are examined with real datasets to evaluate their performance, thus each of their strengths and limitations are compared.



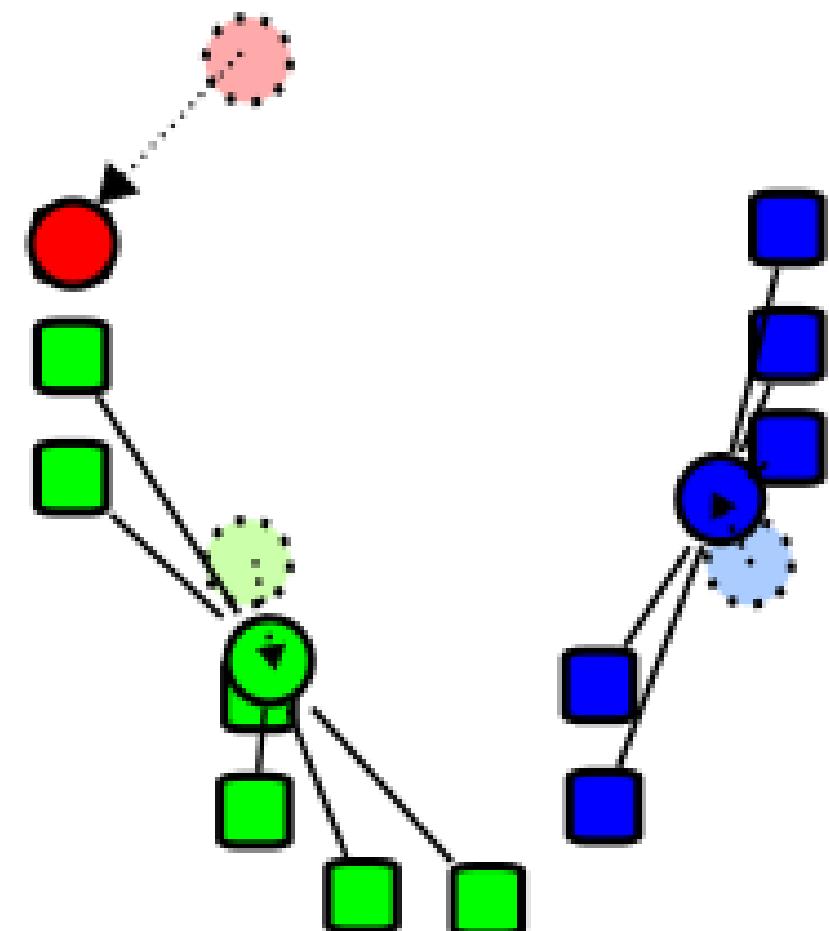
Method

Utilizing different model to recommend a video game to a user of Steam, based on the following:

- K-means clustering
- Gaussian Mixture
- collaborative filitering



K-means clustering



1

INITIALIZATION

Pick a number of clusters and initialize K cluster centroids randomly or using some specific initialization techniques. Each centroid represents a cluster center.

2

ASSIGNMENT

Assign each data point to the nearest cluster centroid based on the Euclidean distance between the data point and the centroids. This step creates clusters by grouping data points based on their proximity to the centroids.

3

UPDATE CENTROIDS

Calculate the mean of all data points assigned to each cluster and update the cluster centroids to the calculated means. This step moves the centroids towards the center of their respective clusters, adjusting their positions to better represent the data points in the cluster.

Then repeat it!

K-means clustering

⋮

K-means clustering is used for partitioning a dataset into distinct groups, or clusters, based on the similarity of data points. It groups similar data points and separates dissimilar ones.

Objective function to minimize:

$$E = \sum_{j=1}^k \sum_{x_n \in C_j} \|x_n - \mu_j\|^2$$

Assignment Step

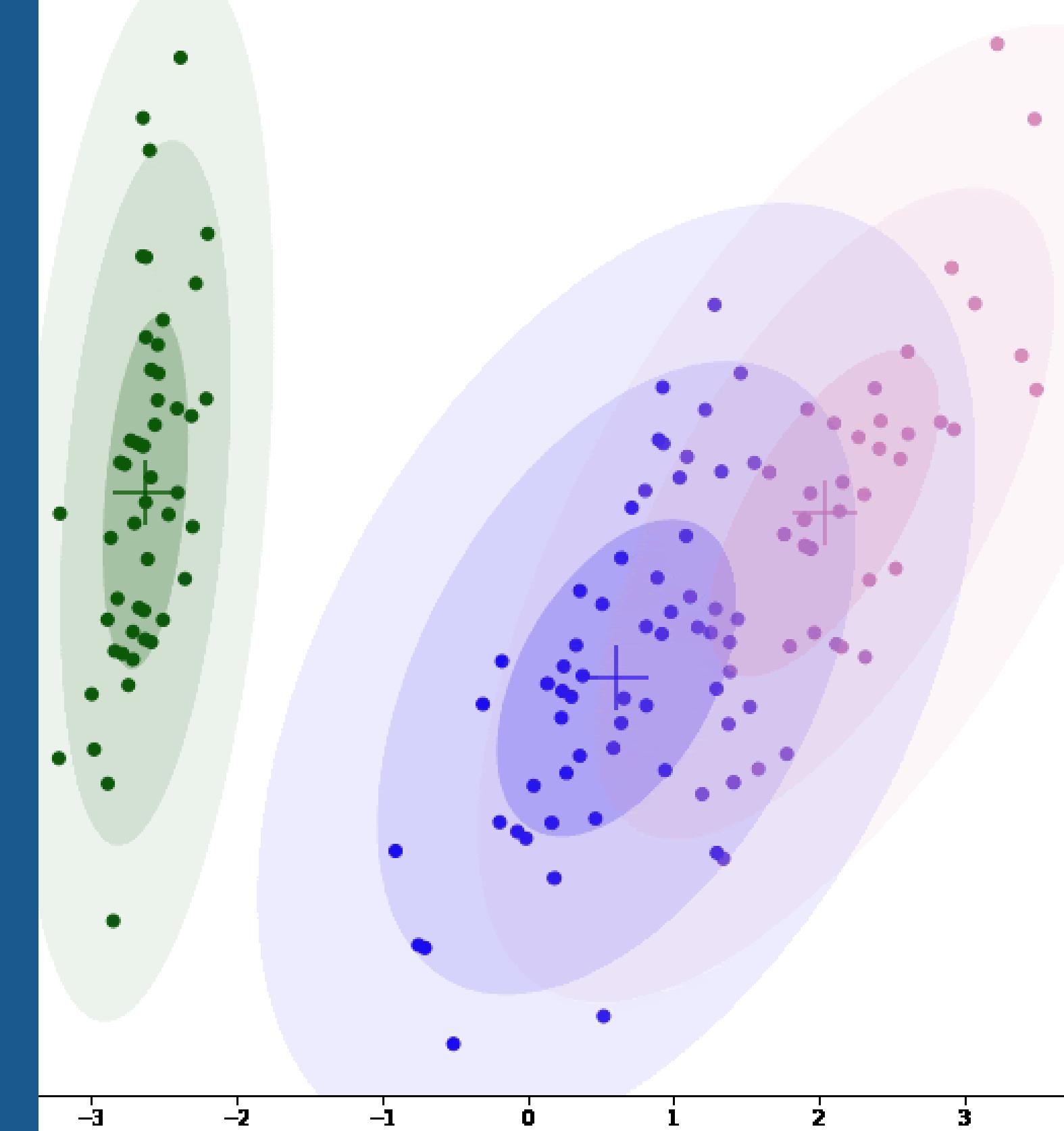
$$C_j = \{x_n : \|x_n - \mu_j\| \leq \|x_n - \mu_i\| \text{ for all } i \neq j\}$$

$$\mu_j = \frac{1}{|C_j|} \sum_{x_n \in C_j} x_n$$

Update Step

Guassian Mixture

A Gaussian Mixture Model is used for representing and describing data that is assumed to be generated from a mixture of several Gaussian distributions. It's a probabilistic model that combines multiple Gaussian (normal) distributions to approximate the underlying distribution of the data.



Normal distributions

Probability density

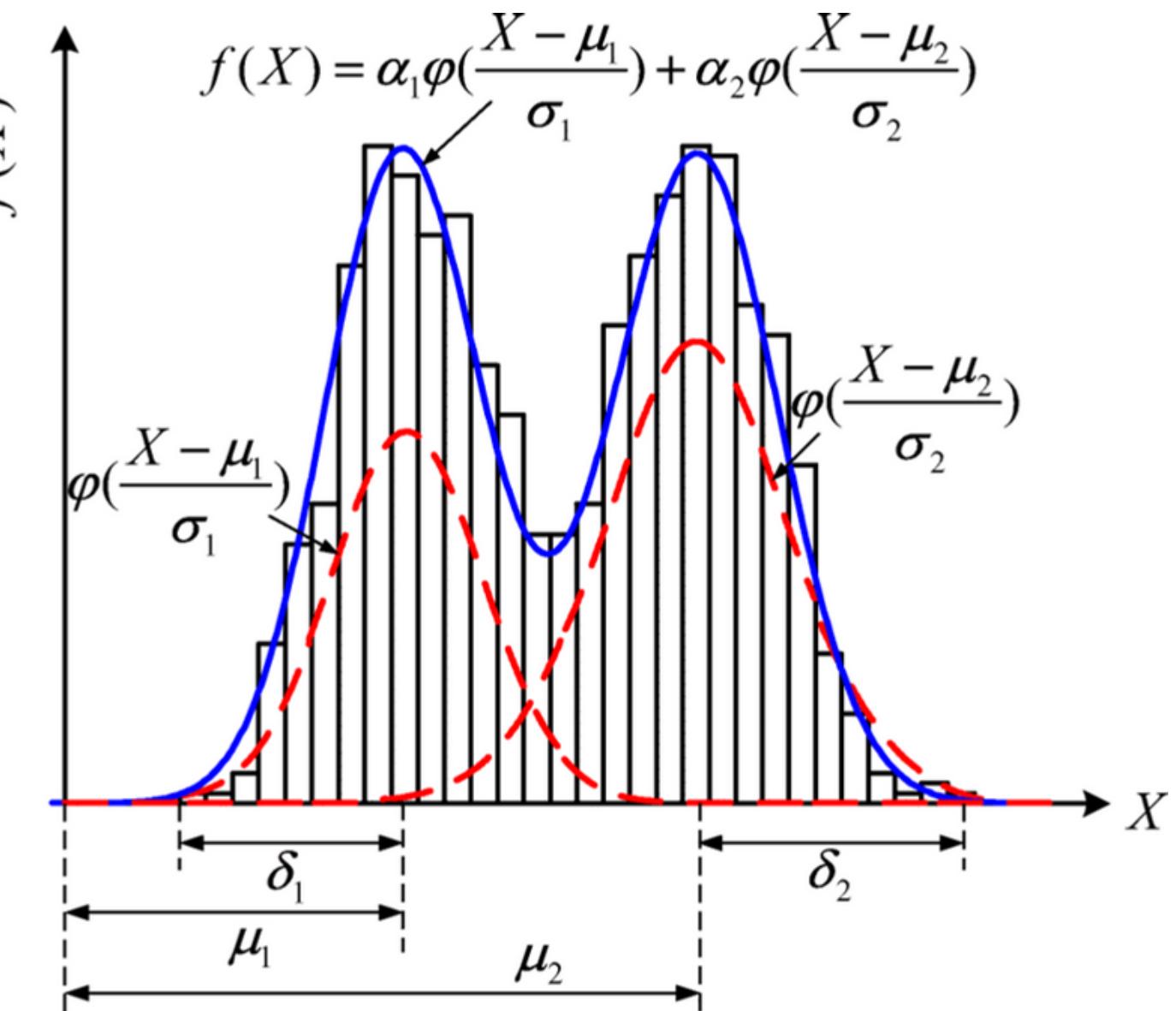
0.8
0.6
0.4
0.2
0.0

-3 -1 1 3

z-score

— A — B — C — D — E

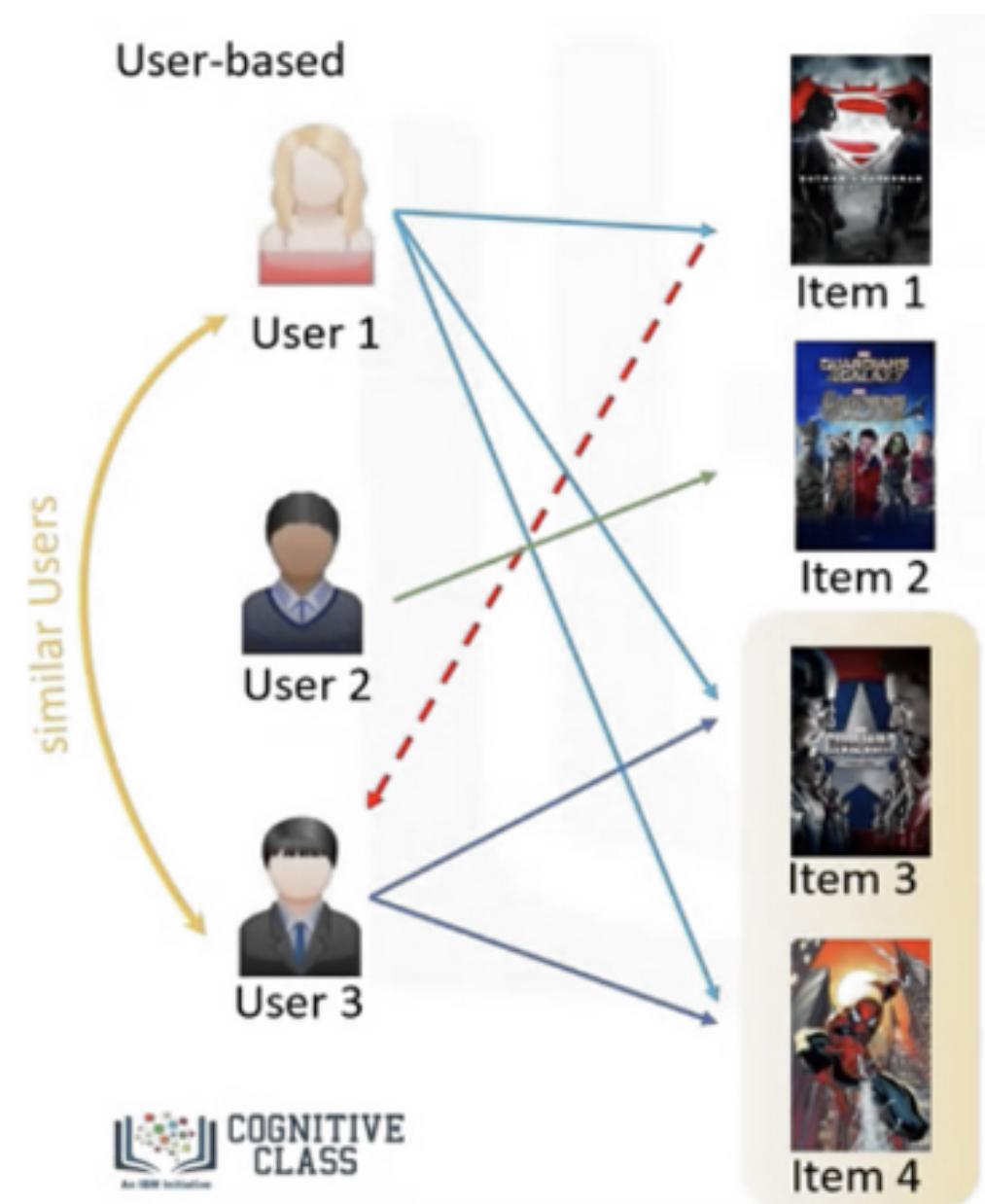
$$f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



Normal distribution, is a continuous probability distribution that is widely used to model and describe random variables. It is characterized by its bell-shaped curve, which is symmetric around its mean.

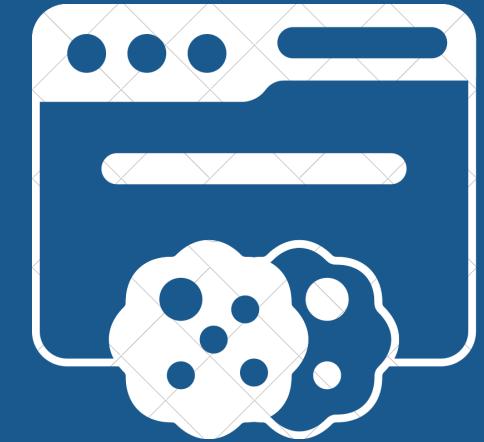
In this case, we use the probability density function to deal with the cluster, It describes the likelihood of a continuous random variable taking on a specific value or falling within a certain range

Collaborative filtering



$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

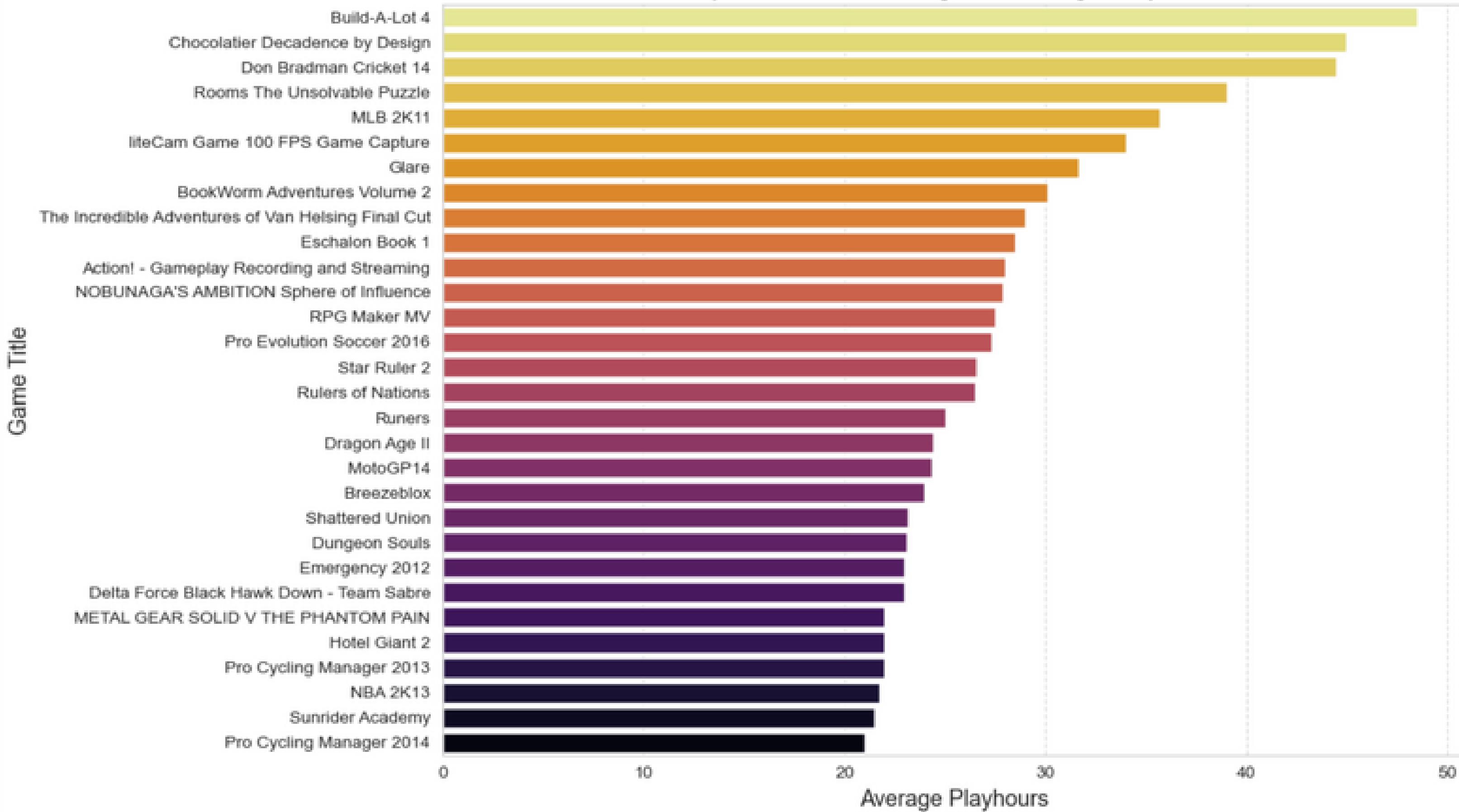
Collaborative Filtering is a technique commonly used in recommendation systems and information filtering to predict a user's preferences or interests by leveraging the preferences and behaviors of a group of users. The fundamental idea behind collaborative filtering is that if users A and B have similar preferences or interests in the past, then the items that A likes might also be liked by B in the future.



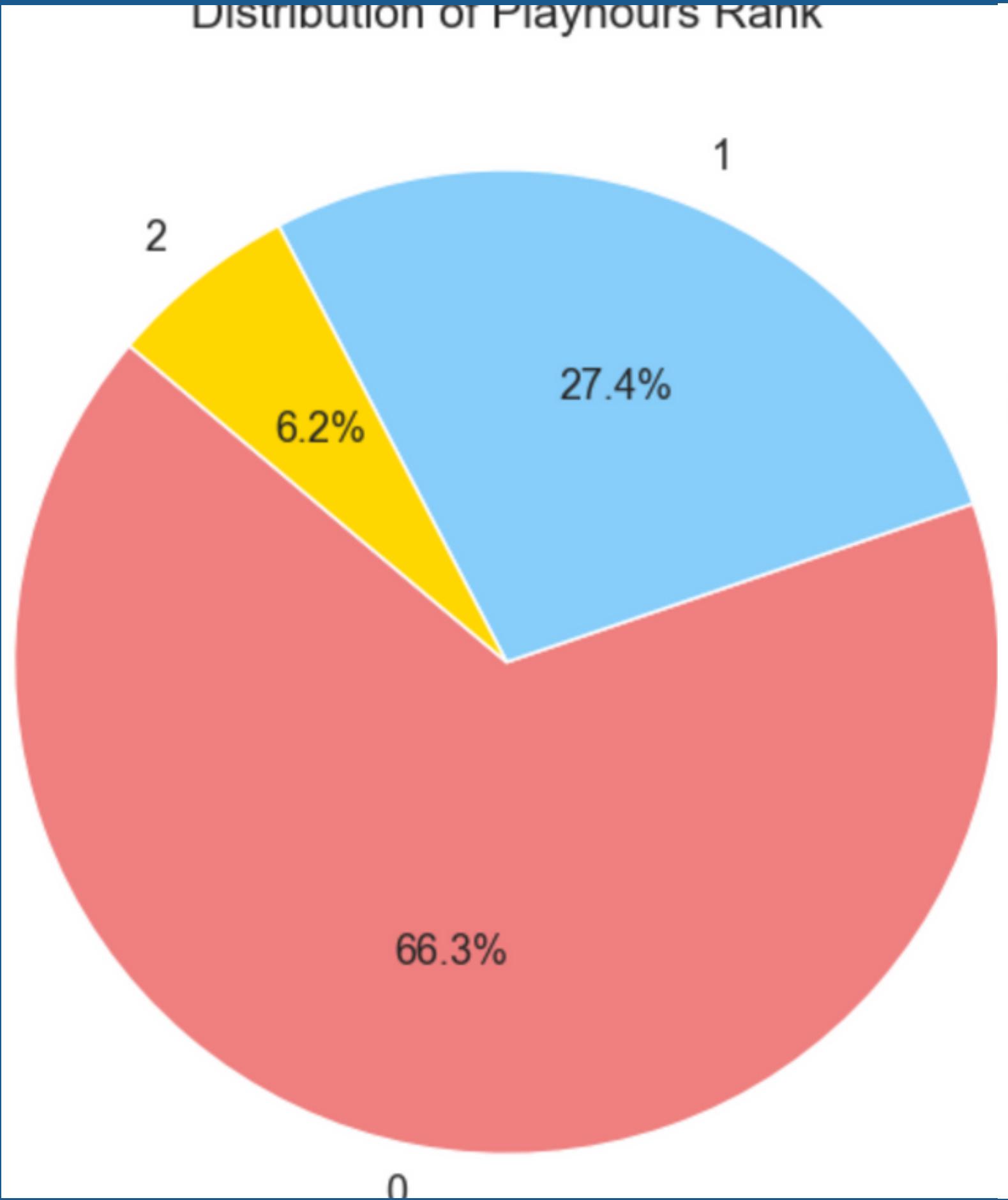
Module presenting



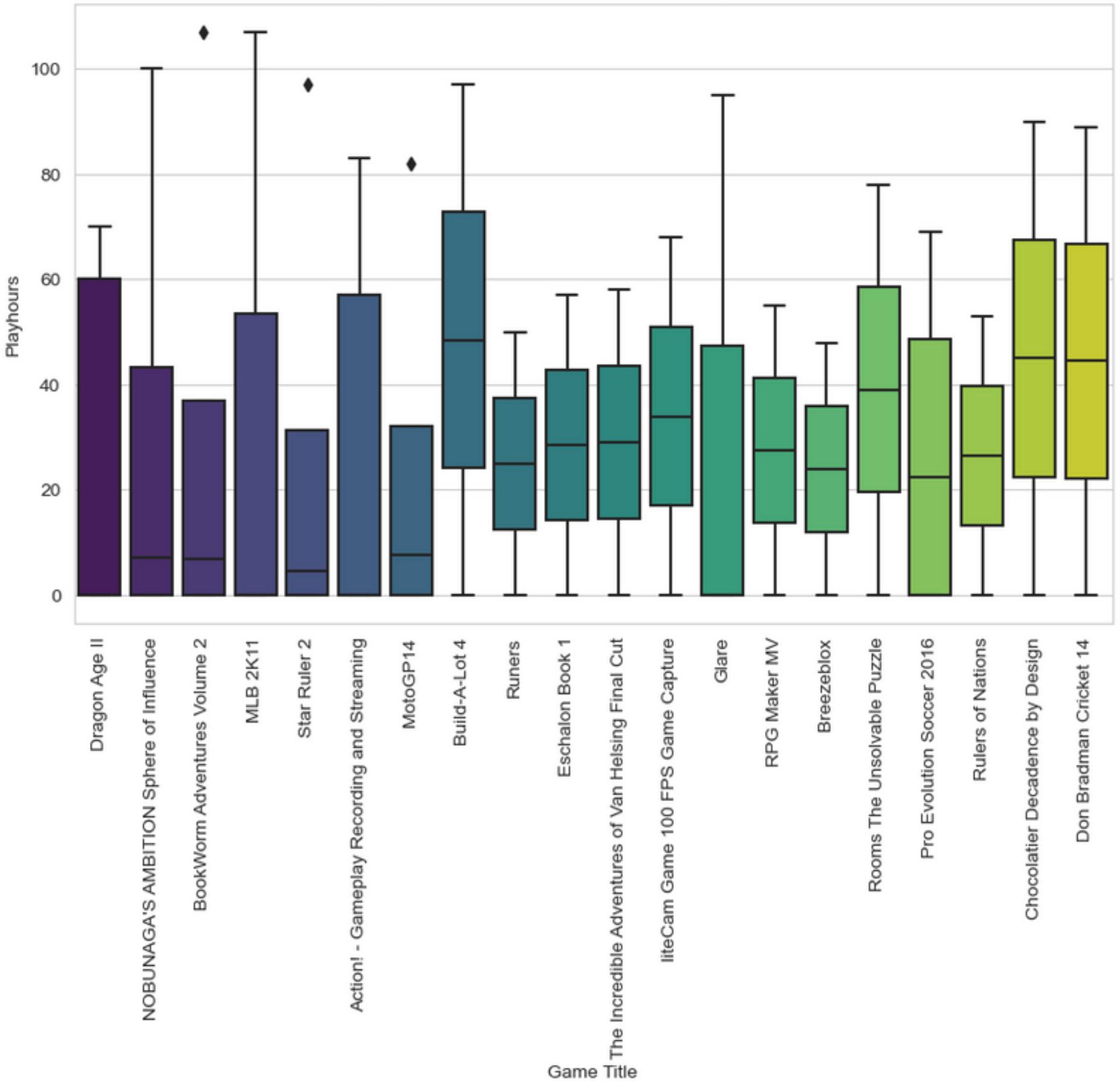
Top 30 Games with Highest Average Playhours



Distribution of Playhours Rank



Playhours Distribution by Game for Top 20 Games



Result

•

•

•

```
*****  
Here are top_10 games recommended for user:54293428  
1:Alien Swarm  
2:ACE - Arena Cyber Evolution  
3:Total War ATTILA  
4:Total War ROME II - Emperor Edition  
5:RAGE  
6:Mafia II  
7:60 Seconds!  
8:Star Wars - Battlefront II  
9:Alan Wake  
10:Arma 2 DayZ Mod  
*****
```

KMeans_Recommender_System([54293428](#),n_top=10)

```
*****  
Here are top_10 games recommended for user:54293428  
1:Unturned  
2:Dota 2  
3:Sid Meier's Civilization V  
4:Left 4 Dead 2  
5:The Elder Scrolls V Skyrim  
6:Warframe  
7:Half-Life 2 Lost Coast  
8:Robocraft  
9:Heroes & Generals  
10:Half-Life 2 Deathmatch  
*****
```

Result

CF_Recommender_System(54293428, 10)

Here are top_10 games recommended for user:54293428

- 1:Airline Tycoon Deluxe
 - 2:Children of the Nile Alexandria
 - 3:Farming Giant
 - 4:Universal Combat CE
 - 5:Imperium Romanum Gold Edition
 - 6:LEGO The Hobbit
 - 7:Dark Souls Prepare to Die Edition
 - 8:Cities Skylines
 - 9:Delta Force Black Hawk Down
 - 10:Terraria
- *****

Although our version of the recommendation system might not be able to compare to The Steam Recommender. We still achieve the goal to construct a recommender system that could recommend games developed by individuals or small game studios.

Discussion

Unsolved problems in our recommender

- The cold start

Optimization for our recommender

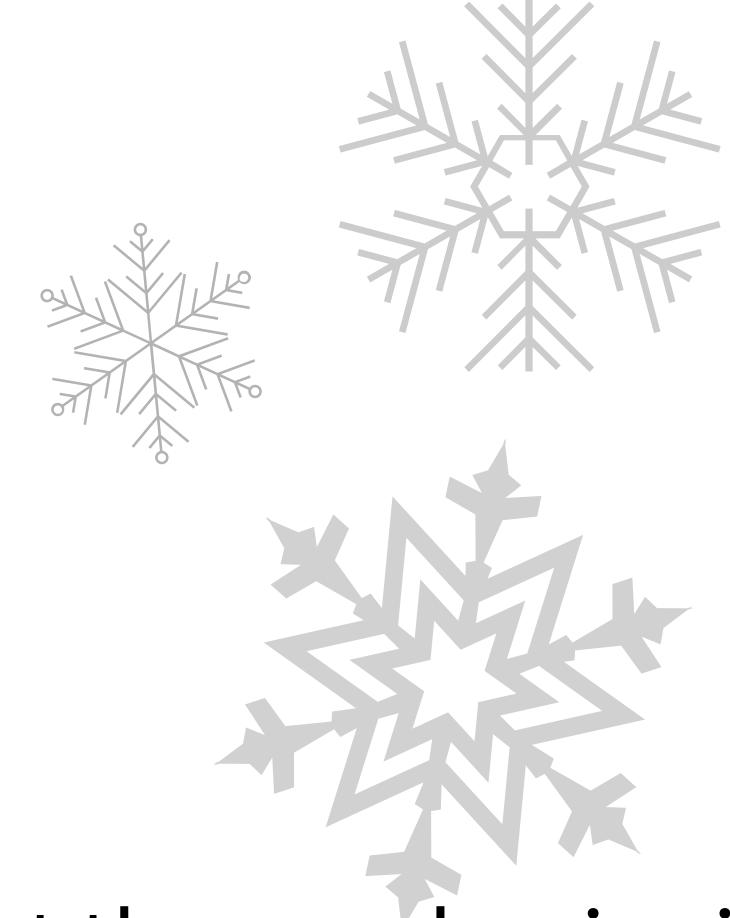
- Temporal Dynamics
- hybrid version

Discussion

The cold start

For new customers :

- We can randomly recommend a few games at the very beginning to see the preference of individuals.
- Assume customers are willing to share their activities on social platform, we can also put it into usage.



Discussion

Temporal Dynamic

Customers' behavior might be affected by current season, therefore we thought it should be a good idea to put some corresponding adjustments to follow the trends.

Ex: selling DLCs(downloadable content)



Conclusion

The first model(Kmeans) is similar to the second one(GMM), which pretty much fail to meet the standard of ours.

Fortunately, the last one(item-based collaborative filtering) successfully made a difference. It features a more objective point of view of recommending(eliminate distraction).

Conclusion

Here are top_10 games recommended for user:54293428

- 1:Alien Swarm
- 2:ACE - Arena Cyber Evolution
- 3:Total War ATTILA
- 4:Total War ROME II - Emperor Edition
- 5:RAGE
- 6:Mafia II
- 7:60 Seconds!
- 8:Star Wars - Battlefront II
- 9:Alan Wake
- 10:Arma 2 DayZ Mod

different

alike

...

KMeans_Recommender_System(54293428,n_top=10)

Here are top_10 games recommended for user:54293428
1:Unturned
2:Dota 2
3:Sid Meier's Civilization V
4:Left 4 Dead 2
5:The Elder Scrolls V Skyrim
6:Warframe
7:Half-Life 2 Lost Coast
8:Robocraft
9:Heroes & Generals
10:Half-Life 2 Deathmatch

different

Here are top_10 games recommended for user:54293428

- 1:Airline Tycoon Deluxe
- 2:Children of the Nile Alexandria
- 3:Farming Giant
- 4:Universal Combat CE
- 5:Imperium Romanum Gold Edition
- 6:LEGO The Hobbit
- 7:Dark Souls Prepare to Die Edition
- 8:Cities Skylines
- 9:Delta Force Black Hawk Down
- 10:Terraria



Thank you for listening