



Steam Playtime Recommendation System

By me, Anthony Mansion

Current Rec Sys State:

As there are 130 million users signed up to Steam and 73,000 games, it's hard to recommend the perfect game to everyone.

On the Steam community tab and other sites like Reddit, there seems to be confusion, disinterest, or maybe even frustration in users being recommend or clicking to look for games.

To improve user engagement, Steam added game playtime to user data to more accurately predict which games one would buy.

Also added playtime because it's more personal than popularity/ratings/genre recommendations.

Agenda

Analysis



Methods



Recommendation System



Future Analysis

Data Analysis

| | User_ID | Videogame | Behavior | Hours_Played |
|--------|-----------|----------------------------|----------|--------------|
| 0 | 151603712 | The Elder Scrolls V Skyrim | purchase | 1.0 |
| 1 | 151603712 | The Elder Scrolls V Skyrim | play | 273.0 |
| 2 | 151603712 | Fallout 4 | purchase | 1.0 |
| 3 | 151603712 | Fallout 4 | play | 87.0 |
| 4 | 151603712 | Spore | purchase | 1.0 |
| ... | ... | ... | ... | ... |
| 199995 | 128470551 | Titan Souls | play | 1.5 |
| 199996 | 128470551 | Grand Theft Auto Vice City | purchase | 1.0 |
| 199997 | 128470551 | Grand Theft Auto Vice City | play | 1.5 |
| 199998 | 128470551 | RUSH | purchase | 1.0 |
| 199999 | 128470551 | RUSH | play | 1.4 |

Perfectly Clean!

- Users: 12.4k
- Videogames: 5,155
- Purchases: 129.5k
- Plays: 70.4k

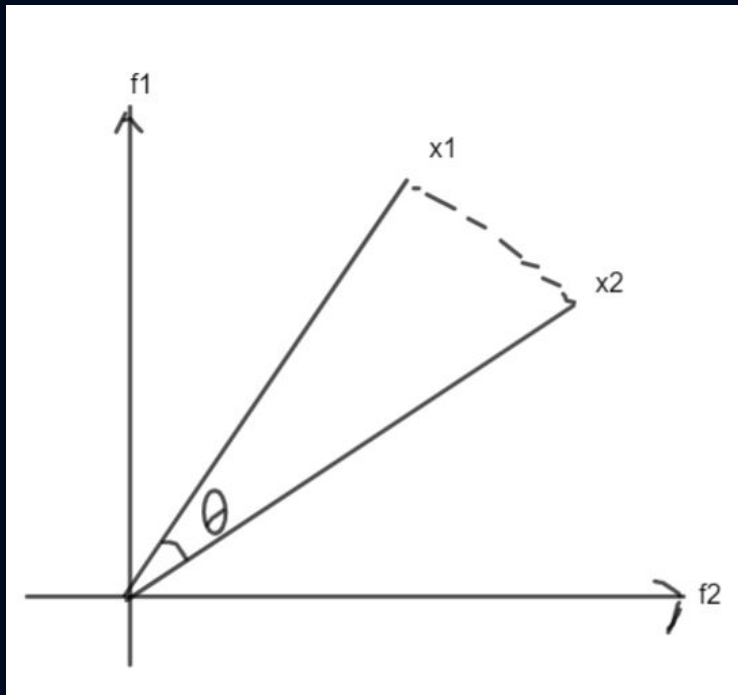
Outliers -

- Highest purchase count: 1k games
- Highest playtime sum: 11.7k hours
- Most played game: 980k hours, 2nd: 322k
- Most frequent game count: 9.6k, 2nd: 4.8k

Random -

- 572 users purchased games, never played them
- 1.8k instances of games played \leq 6 minutes
- 848 games played 6 minutes or less
- 655 games appear once
- Some rows aren't even games

Cosine Similarity



$$\text{cosinesimilarity}(x1, x2) = \cos\theta$$

Imagine two arrows on a graph arrows pointing in different directions. Cosine similarity is the measurement of how similar these 2 arrows are. The smaller the angle, the more similar.

Using cosine similarity, the program was able to compare many vectors of all the videogames, users, and their playtime in a multidimensional space to recommend the 10 most similar games.

User 5250's Recommendations

```
In [30]: # Shows the index of the 10 recommended games from the list as the titles of the 10 recommended games
recommendations(Playtimes.index[user_index])
```

```
['Empire Total War',
 'Stronghold Crusader 2',
 'FIFA Manager 11',
 'Grand Theft Auto V',
 'Homefront',
 'Sid Meier's Civilization V',
 'Total War ROME II - Emperor Edition',
 'RAGE',
 'Call of Duty Modern Warfare 2 - Multiplayer',
 'Ryse Son of Rome',
 'Company of Heroes 2']
```

```
In [31]: # This is the original data on games the random user has played
SteamVGs[SteamVGs["User_ID"]==Playtimes.columns[user_index]]
```

| | User_ID | Videogame | Behavior | Hours_Played |
|-------|---------|--------------------------|----------|--------------|
| 65423 | 5250 | Cities Skylines | purchase | 1.0 |
| 65424 | 5250 | Cities Skylines | play | 144.0 |
| 65425 | 5250 | Deus Ex Human Revolution | purchase | 1.0 |
| 65426 | 5250 | Deus Ex Human Revolution | play | 62.0 |
| 65427 | 5250 | Portal 2 | purchase | 1.0 |
| 65428 | 5250 | Portal 2 | play | 13.6 |
| 65429 | 5250 | Alien Swarm | purchase | 1.0 |
| 65430 | 5250 | Alien Swarm | play | 4.9 |
| 65431 | 5250 | Team Fortress 2 | purchase | 1.0 |
| 65432 | 5250 | Team Fortress 2 | play | 0.8 |
| 65433 | 5250 | Dota 2 | purchase | 1.0 |
| 65434 | 5250 | Dota 2 | play | 0.2 |
| 65435 | 5250 | Counter-Strike | purchase | 1.0 |
| 65436 | 5250 | Counter-Strike Source | purchase | 1.0 |

User 1950243's Recommendations

```
In [30]: # Shows the index of the 10 recommended games from the list as the titles of the 10 recommended games  
recommendations(Playtimes.index[user_index])
```

```
['Dota 2',  
 'Call of Duty Modern Warfare 3',  
 'Spec Ops The Line',  
 'Sniper Elite 3',  
 'Sniper Elite V2',  
 'Counter-Strike Global Offensive',  
 'Team Fortress 2',  
 'Borderlands',  
 'Sid Meier's Civilization V',  
 'Fallout 4',  
 'Counter-Strike Source']
```

```
In [31]: # This is the original data on games the random user has played  
SteamVGs[SteamVGs["User_ID"]==Playtimes.columns[user_index]]
```

| | User_ID | Videogame | Behavior | Hours_Played |
|-------|----------|-----------------|----------|--------------|
| 39858 | 98102518 | Team Fortress 2 | purchase | 1.0 |
| 39859 | 98102518 | Team Fortress 2 | play | 2.7 |

Limitations/Future Analysis Flip

- Categorize games by length
- More playtime specifics
- Videogame publisher type
- Don't recommend purchased vgs



“Questions?” 🧐



STEAM®

