

# Steam Game Recommender



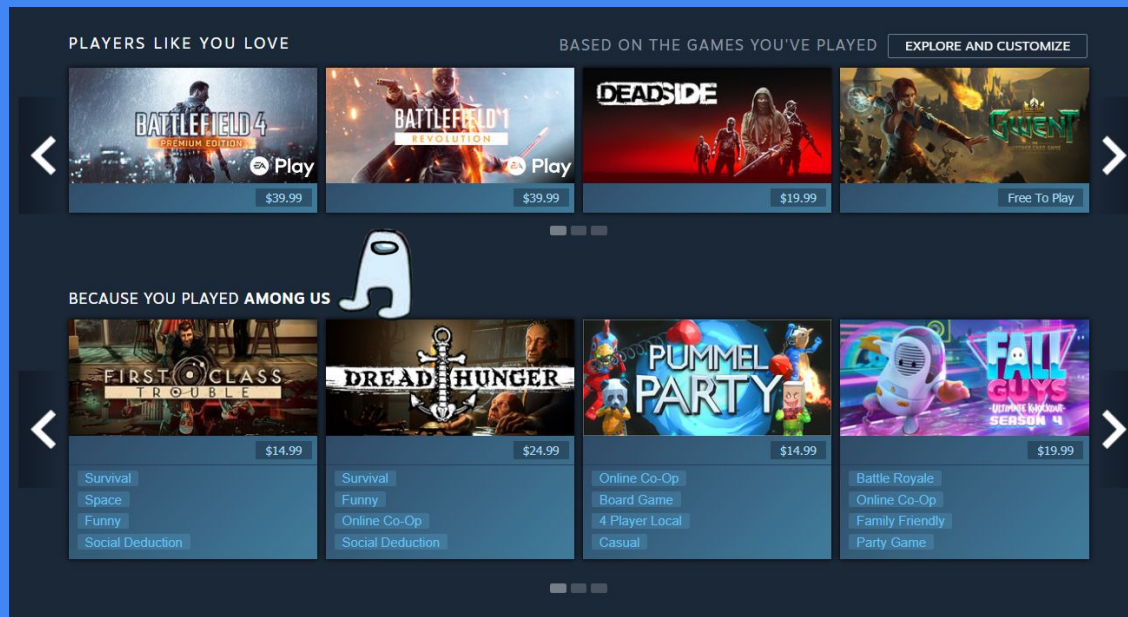
**Mark Sorial, Sam Sukendro, Ari Choe, Coco Gong**

# Problem

Steam is one of the biggest video game distribution services, designed with the purpose of providing users with access to games across multiple computers.

However, finding a good game is hard and time-consuming... so we decided to create a recommender system :)

# What is a game recommender?



Based on an input game, the system will recommend similar content

# Dataset



- Steam game data from Kaggle
- About ~3000 (filtered) games in our dataset
- About ~300 individual tags to work with

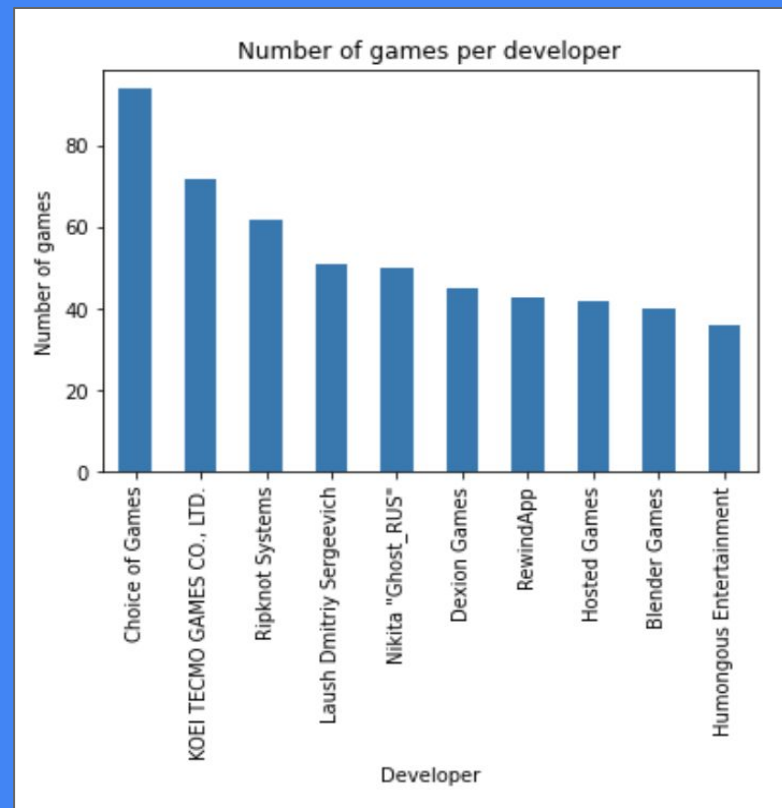
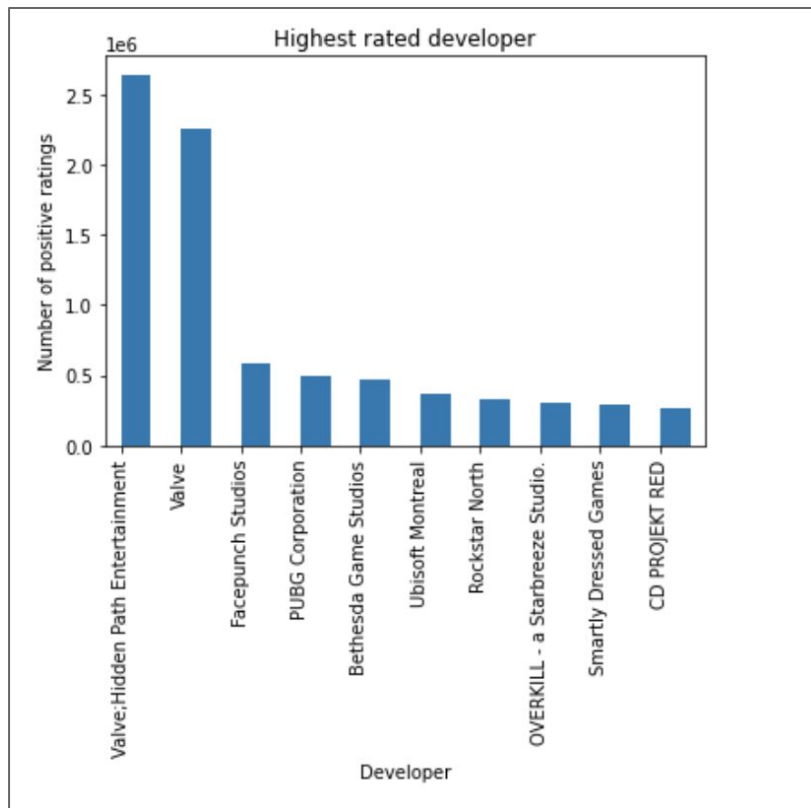
## Example Data

appid int64	name object	release_date object	categories object	genres object	steamspy_tags object
10 - 1852070	Counter-Strike ..... 0%	2016-04-05 ..... 0.4%	['Single-player'] ..... 6.4%	['Action'] ..... 9.1%	['Casual', 'Indie'] ..... 1.2%
	Team Fortress C_ ..... 0%	2007-08-03 ..... 0.3%	['Full controller _ ..... 6.1%	['Action', 'Ind_ ..... 5.5%	['Adventure', 'Ind_ ..... 0.9%
	3237 others ..... 99.9%	1696 others ..... 99.3%	1237 others ..... 87.5%	496 others ..... 85.4%	1746 others ..... 98%

0	10	Counter-Strike	2000-11-01	['Local Multi-Player', 'Multi-player', 'Online Multi-Player', 'Valve ...	['Action']	['Action', 'FPS', 'Multiplayer']
---	----	----------------	------------	--	------------	-------------------------------------

# Interesting Graphs



# Game Data/Issued Data

Data column	<code>'appid', 'name', 'release_date', 'categories, genres', 'steamspy_tags', 'english', 'positive_ratings', 'negative_ratings', 'average_playtime'</code>
Dropped data	<code>'developer', 'publisher', 'platforms', 'required_age', 'achievements', 'owners', 'english', 'price', 'average_playtime', 'positive_ratings', 'negative_ratings', 'release_date'</code>
	<code>postive_rating &lt; 500, english = 0</code>

To clean up our data, we dropped irrelevant information and filtered out lesser known games.

# Processing Data

## Categorical Data



## Binarized Data

['Local Multi-Player', 'Multi-player', 'Online Multi-Player', 'Valve ...	['Action']	['Action', 'FPS', 'Multiplayer']
['Local Multi-Player', 'Multi-player', 'Online Multi-Player', 'Valve ...	['Action']	['Action', 'FPS', 'Multiplayer']
['Multi-player', 'Valve Anti-Cheat enabled']	['Action']	['FPS', 'Multiplayer', 'World War II']
['Local Multi-Player', 'Multi-player', 'Online Multi-Player', 'Valve ...	['Action']	['Action', 'FPS', 'Multiplayer']
['Multi-player', 'Single-player', 'Valve Anti-Cheat enabled']	['Action']	['Action', 'FPS', 'Sci-fi']

['Cross-Platform Multiplayer', 'In-App Purchases', 'Online ...	['Early Access', 'Free to Play', 'Strategy']	['Early Access', 'Free to Play', 'MOBA']
['Partial Controller Support', 'Single-player', 'Steam ...	['Indie']	['Bullet Hell', 'Faith', 'Indie']
['Single-player', 'Steam Cloud']	['Casual', 'Indie', 'Simulation']	['Casual', 'Dating Sim', 'Psychological Horror']
['Single-player', 'Steam Achievements', 'Steam Leaderboards']	['Casual', 'Indie', 'Strategy']	['City Builder', 'Indie', 'Strategy']
['Single-player']	['Adventure', 'Free to Play', 'Indie']	['Adventure', 'Free to Play', 'Indie']

1980s	2D	2D Fighter	3D	3D Platformer	4X	6DOF	Action	Action RPG	Adventure	Agriculture
0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0
...	...	...	...	...	...	...	...	...	...	...
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0

In-App Purchases	Includes Source SDK	Includes level editor	Local Co-op	Local Multi-Player	MMO	Mods
0	0	0	0	1	0	0
0	0	0	0	1	0	0
0	0	0	0	0	0	0
0	0	0	0	1	0	0
0	0	0	0	0	0	0
...	...	...	...	...	...	...
1	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

# Methods

After dropping/filtering extra info and binarizing the relevant data...

## KNN

The Euclidean distance between the user's input game and "neighboring" games are calculated.

Shorter distance = more similar

## Cosine Similarity

In an inner product space, the cosine of the angle between two vectors are measured.

Smaller angles = more similar

\*Note: the similarity score included with this model is NOT angle measure.

Higher similarity score = more similar



# Models/Modified Parameters



Some extra algorithms to maximize performance...

## KNN

Algorithm:

NearestNeighbors from  
sklearn.neighbors

Specifics:

n\_neighbors = 6  
algorithm = 'ball\_tree'

## Cosine Similarity

Algorithm:

cosine\_similarity from  
sklearn.metrics.pairwise

Specifics:

CountVectorizer from  
sklearn.feature\_extraction.text

# Results (KNN Model)

Selected Game: The Elder Scrolls V: Skyrim

Tags: Fantasy, Open World, RPG

Similar Names:

The Elder Scrolls V: Skyrim 414

The Elder Scrolls V: Skyrim Special Edition 2461

The Elder Scrolls V: Skyrim VR 2839

Recommendations:

Kingdoms of Amalur: Reckoning ['Fantasy', 'Open World', 'RPG']

FINAL FANTASY IV ['Fantasy', 'JRPG', 'RPG']

FINAL FANTASY XIII-2 ['Fantasy', 'JRPG', 'RPG']

Pillars of Eternity II: Deadfire ['Fantasy', 'Isometric', 'RPG']

The Elder Scrolls III: Morrowind® Game of the Year Edition ['Fantasy', 'Open World', 'RPG']

Selected Game: DayZ

Tags: Open World, Survival, Zombies

Similar Names:

DayZ 670

Recommendations:

Reign Of Kings ['Medieval', 'Open World', 'Survival']

Next Day: Survival ['Multiplayer', 'Open World', 'Survival']

Dying Light ['Open World', 'Survival', 'Zombies']

Nether: Resurrected ['Action', 'Open World', 'Survival']

SURVIVAL ['Early Access', 'Open World', 'Survival']

Selected Game: Counter-Strike

Tags: Action, FPS, Multiplayer

Similar Names:

Counter-Strike 0

Counter-Strike: Condition Zero 7

Counter-Strike: Source 10

Counter-Strike: Global Offensive 25

Counter-Strike Nexon: Zombies 1175

Recommendations:

Deathmatch Classic ['Action', 'FPS', 'Multiplayer']

Counter-Strike ['Action', 'FPS', 'Multiplayer']

Ricochet ['Action', 'FPS', 'Multiplayer']

Half-Life Deathmatch: Source ['Action', 'FPS', 'Multiplayer']

Counter-Strike: Condition Zero ['Action', 'FPS', 'Multiplayer']

Selected Game: Team Fortress 2

Tags: FPS, Free to Play, Multiplayer

Similar Names:

Team Fortress 2 19

Recommendations:

Paladins® ['FPS', 'Free to Play', 'Multiplayer']

Block N Load ['Action', 'Free to Play', 'Multiplayer']

Counter-Strike: Global Offensive ['FPS', 'Multiplayer', 'Shooter']

Blockstorm ['Action', 'FPS', 'Multiplayer']

Batla ['FPS', 'Free to Play', 'Multiplayer']

# Results (Cosine Similarity)

## Cosine Similarity Model

Selected Game: The Elder Scrolls V: Skyrim

### Similar Names:

The Elder Scrolls V: Skyrim 414  
The Elder Scrolls V: Skyrim Special Edition 2461  
The Elder Scrolls V: Skyrim VR 2839

### Reccomened Games:

The Elder Scrolls V: Skyrim Special Edition -- Similarity: 0.8003267306650416  
ELEX -- Similarity: 0.7559289460184543  
Kingdoms of Amalur: Reckoning™ -- Similarity: 0.7302967433402214  
My Time At Portia -- Similarity: 0.7302967433402214  
FINAL FANTASY® XIII-2 -- Similarity: 0.7222222222222224

## Cosine Similarity Model

Selected Game: DayZ

### Similar Names:

DayZ 670

### Reccomened Games:

Half-Life -- Similarity: 0.727532837268552  
Miscreated -- Similarity: 0.7275328372685519  
Out of Reach -- Similarity: 0.7110571291368637  
Dying Light -- Similarity: 0.7010681840159741  
Next Day: Survival -- Similarity: 0.6982972487551757

## Cosine Similarity Model

Selected Game: Counter-Strike

### Similar Names:

Counter-Strike 0  
Counter-Strike: Condition Zero 7  
Counter-Strike: Source 10  
Counter-Strike: Global Offensive 25  
Counter-Strike Nexon: Zombies 1175

### Reccomened Games:

Deathmatch Classic -- Similarity: 0.8947368421052628  
Ricochet -- Similarity: 0.8902638145194516  
Team Fortress Classic -- Similarity: 0.8720815992723809  
Counter-Strike: Condition Zero -- Similarity: 0.8346223261119858  
Half-Life -- Similarity: 0.7648661601319408

## Cosine Similarity Model

Selected Game: Team Fortress 2

### Similar Names:

Team Fortress 2 19

### Reccomened Games:

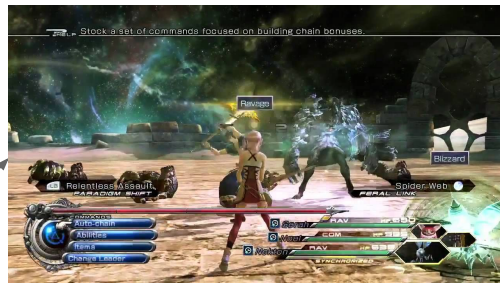
Nuclear Dawn -- Similarity: 0.7392493243604371  
Insurgency -- Similarity: 0.733235575106767  
BRAIN / OUT -- Similarity: 0.7275476020141096  
Depth -- Similarity: 0.7099522928088307  
No More Room in Hell -- Similarity: 0.7003892132633234

# Results (KNN)

Top 3 Recommended games  
(ignoring the other final fantasy)



User input: **Skyrim**  
(Open-World RPG)



**Final Fantasy XII-2**  
(Strategy JRPG)



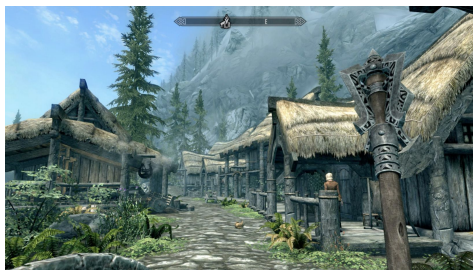
**Kingdoms of Amalur**  
(Open-World RPG)



**Pillars of Eternity II**  
(Isometric RPG Strategy)

# Results (Cosine Similarity)

Top 3 Recommended games  
(ignoring the other version of skyrim)



User input: **Skyrim**  
(Open-World RPG)



**ELEX**  
(Open-world RPG)



**Kingdoms of Amalur**  
(Open-world RPG)



**My Time at Portia**  
(Open-world  
RPG/Farming Sim)

## A Note on Evaluation

This model is based on content recommendation, so there is no definite “ground truth” that the results can be measured against. As a result, conventional metrics like MSE cannot be applied here.





# Conclusion

In terms of model performance...

1. We printed the tags of the input game and the recommendation games
2. After comparison, we found that Cosine Similarity recommendations had tags that matched the input more closely

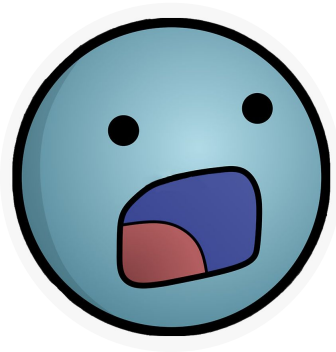
In terms of algorithmic concepts...

1. Cosine similarity handles larger amounts of dimensions better than KNN
2. Results of Cosine Similarity generated quantitative scores, which are more easily measured

**Thus, our cosine similarity model is the better performing model due to better model performances and algorithmic concepts.**

# Limitations

- Due to the lack of ground truth, the models cannot be truly compared
- Steamspy\_tags are actually user created, which may lead to inconsistencies in the data
- We limited our dataset to about 3,000 games to ensure our recommendations had over 500 ratings, but there are actually over 50,000 games in the database





# References

Davis, Nik. "Steam Store Games (Clean Dataset)." *Kaggle*, 12 June 2019, [www.kaggle.com/nikdavis/steam-store-games](https://www.kaggle.com/nikdavis/steam-store-games).

Prabhakaran, Selva. "Cosine Similarity - Understanding the Math and How It Works? (With Python)." *Machine Learning Plus*, 22 Oct. 2018, <https://www.machinelearningplus.com/nlp/cosine-similarity/>.

Harrison, Onel. "Machine Learning Basics with the K-Nearest Neighbors Algorithm." *Medium*, 14 July 2019, <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>.

**THANK YOU**



# NATHAKINISM

**NATHAKIN FOR THE WIN**

**BEST TA**

NATHAKIN FOR THE WIN  
NATHAKIN FOR THE WIN  
NATHAKIN FOR THE WIN  
NATHAKIN FOR THE WIN  
NATHAKIN FOR THE WIN  
NATHAKIN FOR THE WIN  
NATHAKIN FOR THE WIN  
NATHAKIN FOR THE WIN  
NATHAKIN FOR THE WIN  
NATHAKIN FOR THE WIN

We love nathakin very much<33333

THANK YOU SO MUCHH!!