

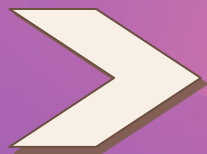


x3

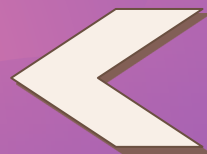
WHAT SHOULD I PLAY?



120



STEAM VIDEO GAME RECOMMENDER SYSTEM



Presented by Leticia Genao



x3

Contents of this Presentation

- 01** • What is Steam?
 - History & Facts
 - Steam's Problem
- 02** • Problem Statement
- 03** • Modeling Goals
- 04** • Steam's Current System
- 05** • Overview of Data
- 06** • Visualizations
 - Top Unique Words
 - Bi-Grams
 - Tri-Grams
- 07** • Approach
- 08** • Recommender Comparison
 - TFIDF on DESC
 - CVEC on DESC
 - CVEC on TAGS
- 09** • Conclusion & Findings
- 10** • Future Improvements
- 11** • Thank You
- 12** • References

Steam's History & Facts

What is Steam?

Founded in 2002 and developed by Valve Corporation, Steam is a digital game distribution platform which offers digital right management, multiplayer gaming, video streaming, and social networking services.

Annual Earnings

Estimated at 100 million sale dollars per year and leading in the digital game publishing industry ahead of its competitors Epic Games, OP.GG, Origin, and Humble Store

Humble Beginnings

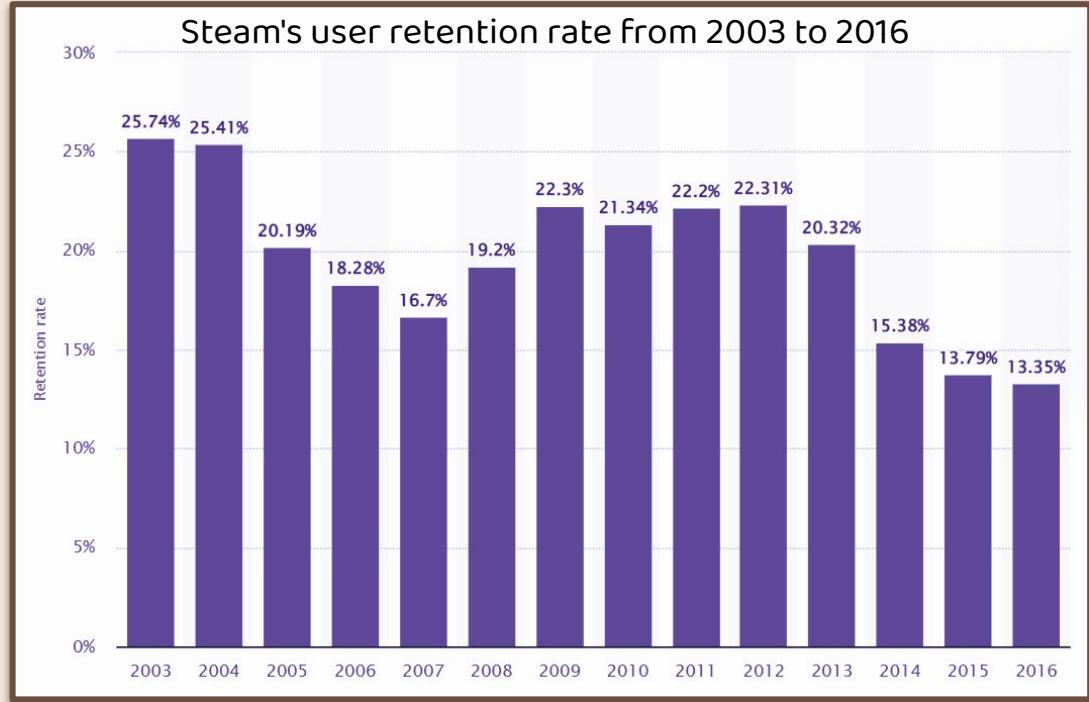
In its very beginnings, Steam was a tiny platform that released only seven games in 2004. Today they have a total of 50K+ games available ranging from Goat Simulator to Grand Theft Auto V. ([Statistia](#))

Average Users

Steam had approximately 120 million monthly active players in 2020, a significant increase from 95 million monthly active users in 2019. ([Statistia](#))

Steam's Problem

As more competitors have popped up over the years, Steam's user retention rate has decreased from 25% to a mere 13%.



Problem Statement

With many video games retailers in the market, the competition to retain customers is highly competitive with multiple platforms available for games to purchase and play through. With various choices to make, keeping customers buying games from your platform is essential.

This project aims to develop a video game recommendation system for Steam users to input their favorite video game title and have similar games suggested in order to keep users from switching to other platforms such as Epic Games, OP.GG, GOG, Humble Store, Itch.io, and Fanatical.

Modeling Goals

- **Test various purely content based recommender systems and identify which has the most relevant recommendations**
- **Identify useful patterns in the text of the dataset**
- **Use recommender system with best cosine similarity scores for system**

Background: Steam's Current System

According to an article by Stephen Totilo, author of Axios Gaming, Steam leans on users' [curation pages](#) and player reviews on the games themselves to raise awareness for games on their platform (Totilo, 2022).

While game reviews and player recommendation are an engaging way for users to submit their opinions and recommend titles to other players, what do you do when someone hasn't reviewed a game you want to push on your platform yet? That's where a recommender system will come in handy.

Steam's [current](#) game recommendation system has limited capabilities. You can only search by tag, age, and popularity.

Overview of Data

01

Description

Kaggle dataset from 2019 containing all 40,833 games available and set to release by 2023.

02

Full Games Only

Removed betas, trials, demos, unreleased games, nulls, non-English, and mis-priced games above \$100.

03

Final Features

Name, developer, popular_tags, game_details, genre, and game_description

04

Final Shape

Rows: 20,969
TFIDF Columns: 70,143
CVEC Columns: 14,899



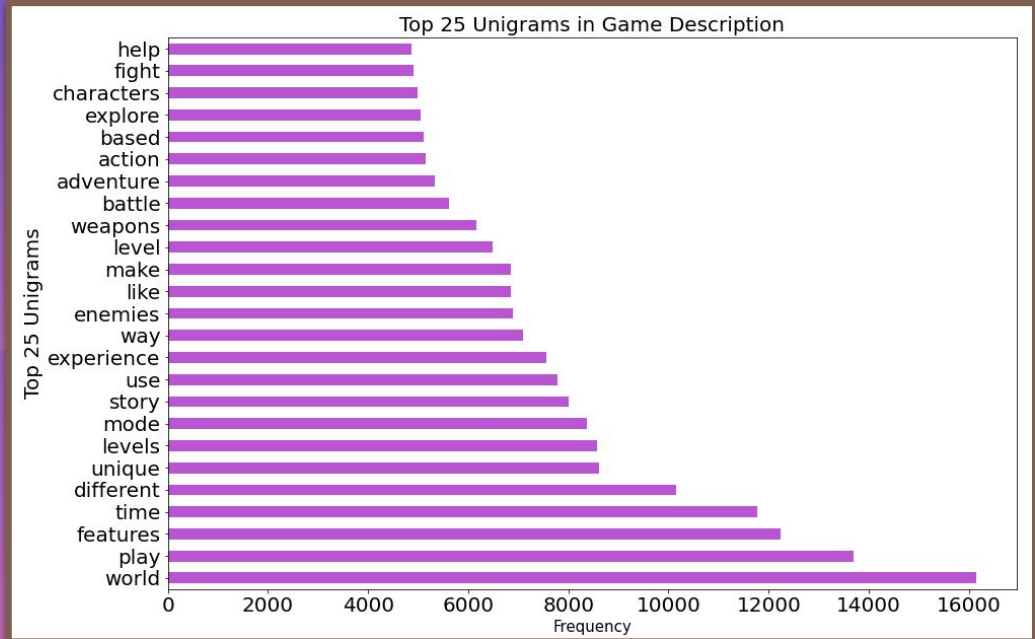
x3

Visualizations: 25 Top Unique Words

Parameters

`min_df = 5, max_df = .90, max_features = 5000, stop_words = MY_STOP`

`my_words=['game', 'games', 'new', 'gameplay', 'players', 'player', 'just']`

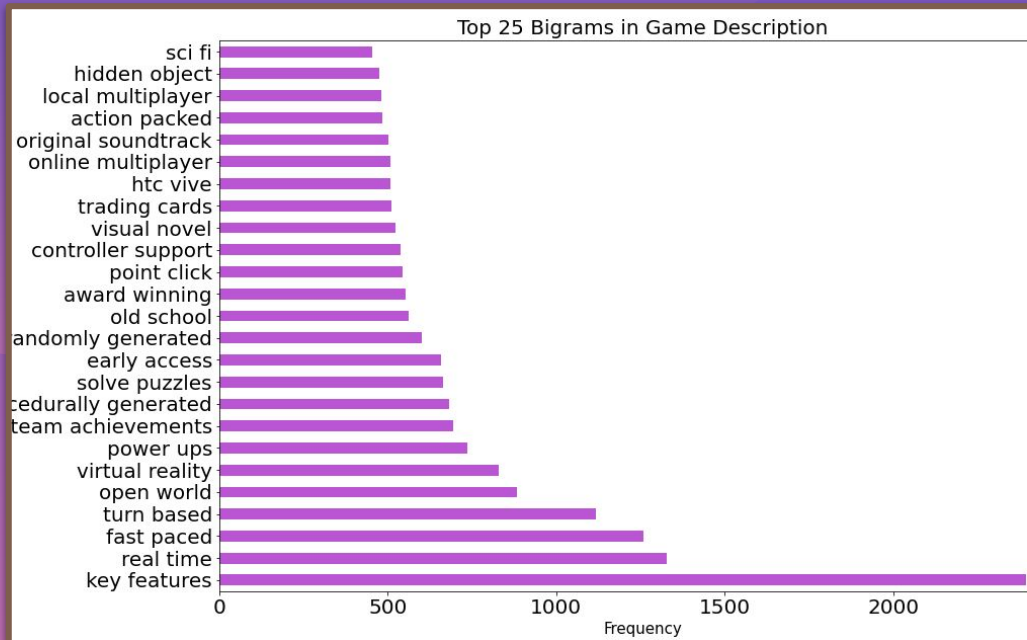


Visualizations: 25 Top Bigrams

Parameters

`min_df = 5, max_df =
.90, max_features = 5000,
stop_words = MY_STOP`

`my_words=['game',
'games', 'new','gameplay',
'players', 'player', 'just']`

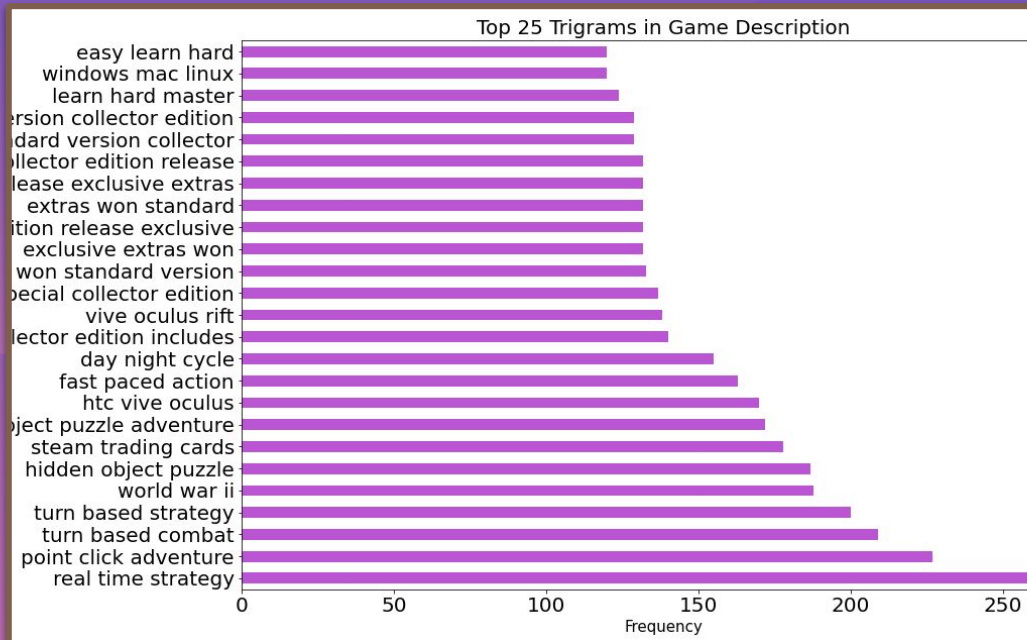


Visualizations: 25 Top Trigrams

Parameters

`min_df = 5, max_df = .90, max_features = 5000, stop_words = MY_STOP`

`my_words=['game', 'games', 'new', 'gameplay', 'players', 'player', 'just']`



Approach

Description Based Recommender System

TFIDF Vectorizer: As TFIDF negates high frequency words we aim to find distinct insightful text using English and custom stopwords. Comparison of cosine similarities on the **text description** with relevant recommendations will be evaluation. Final Shape: (20969, 70143)

Countvectorizer: Conversely, I wanted to test using commonality between descriptions to see how the cosine similarity scores compared to the recommended TFIDF scores.

Tag Based Recommender System

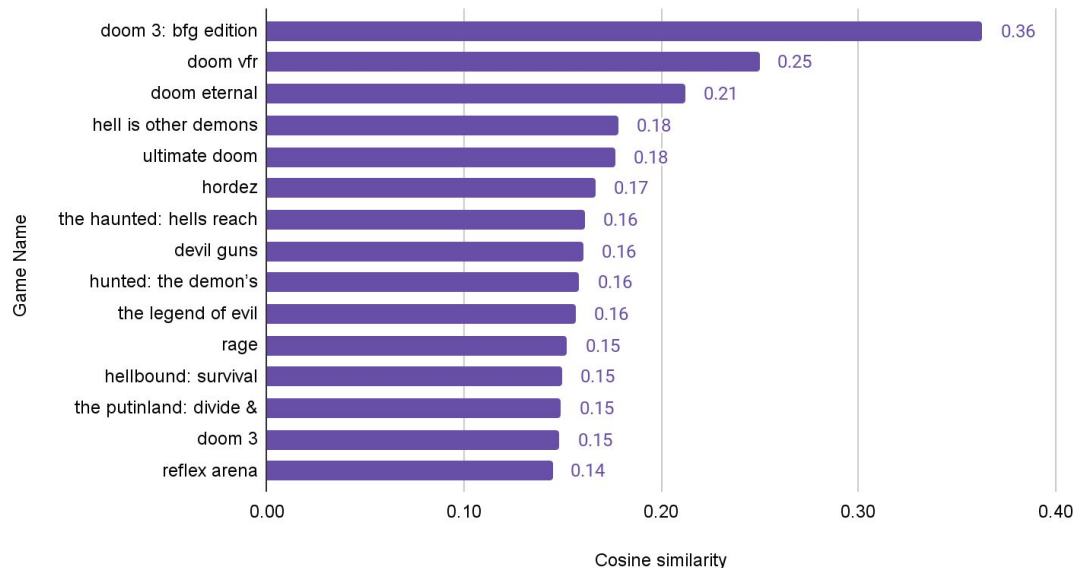
Countvectorizer: Unlike with TFIDF, which negates high frequency words, we desire the emphasis on frequently occurring tags to base the recommendations on. The additional weights of countvectorizer will improve our recommendations so this will be paired with searching over the first 3-5 weighted tags of each feature while combining tags into a metasoup of text testing English and custom stopwords for score improvement.

Model Comparison: TFIDF DESC - Top 15

Top Games

1. Doom 3: BFG Edition
2. Doom VFR
3. Doom Eternal
4. Hell is Other Demons
5. Ultimate Doom

TFIDF Cosine Similarity - Description

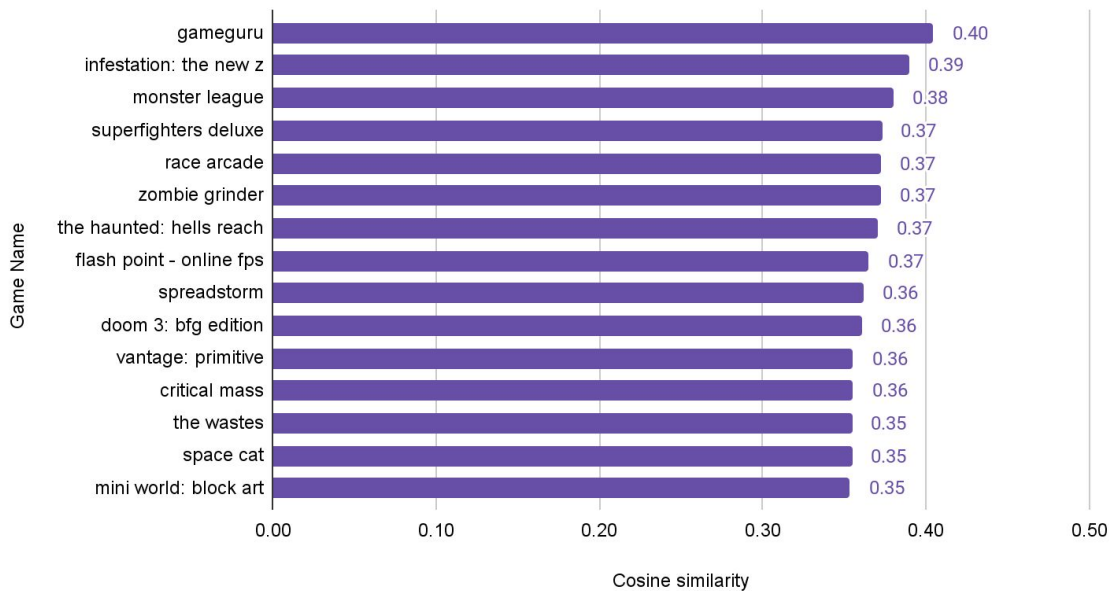


Model Comparison: CVEC DESC - Top 15

Top Games

1. Gameguru
2. Infestation: The New Z
3. Monster League
4. Superfighters
5. Race Arcade

CVEC Cosine Similarity - Description

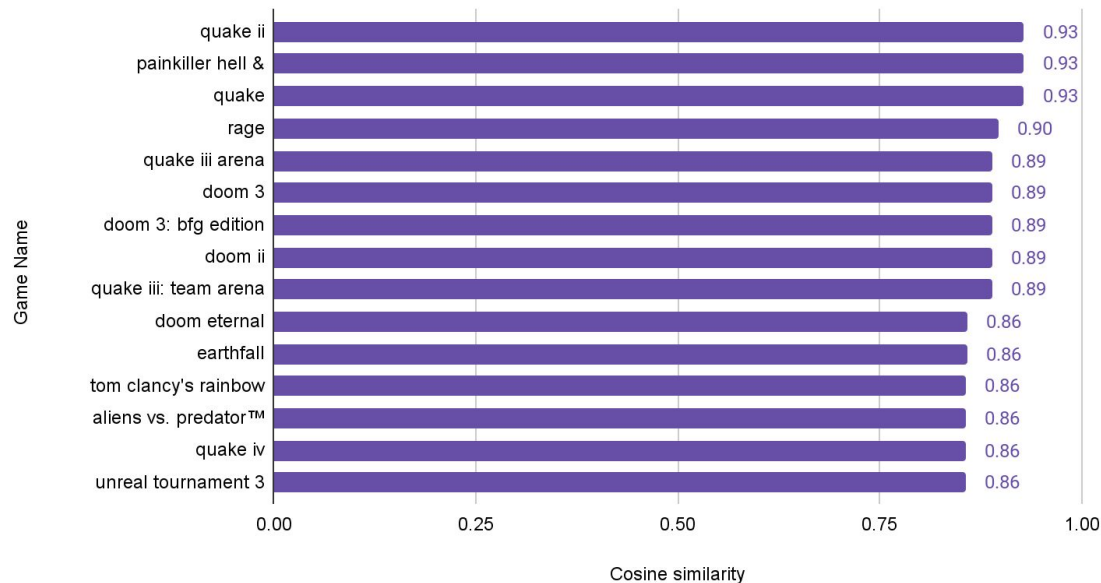


Model Comparison: CVEC TAGS- Top 15

Top Games

1. Quake II
2. Painkiller Hell & Damnation
3. Quake
4. Rage
5. Quake III Arena

CVEC Cosine Similarity - TAGS



Conclusion & Findings



CVEC Wins!

The best recommender system was a countvectorizer using the features of developer, popular_tags, game_details, and genre with a 3 weight tag prioritization and English stop words



TFIDF low scores

TFIDF vectorizer which is known to be a popular metric to use for description based content recommender systems had the lowest cosine scores



CVEC on DESC

The countvec description recommender yielded higher cosine similarity scores than the TFIDF, but more irrelevant recommendations when manually inspected.

Future Considerations

1

Description + Tags

Creating a combo text recommender system could yield better results.

2

More Tags!

Maturity rating, user rating, and steam positive review score would increase the scoring of the system

3

Price tier

Games recommended by price category would aid users wanting games on sale

4

A Collaborative Combination

Combining various scoring, weights, tags, and descriptions could provide a complex and rewarding collaboration recommender system which algorithms like PCA , SVD, etc can be used and properly evaluated.





Thank You For Playing!

Continue?

CREDITS: This presentation template was created by **Slidesgo**,
including icons by **Flaticon** and infographics & images by **Freepik**
Please keep this slide for attribution

Sources

- ALEXANDER ANTONOV, (2019). Steam games complete dataset.
<https://www.kaggle.com/datasets/trolukovich/steam-games-complete-dataset>
- F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh, (2015). Recommendation systems: Principles, methods and evaluation. Egyptian Informatics Journal, Pages 261-273. <https://doi.org/10.1016/j.eij.2015.06.005>
- G. Shani, A. Gunawardana. (2009). MSR-TR-2009-159, Evaluating Recommender Systems
https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/EvaluationMetrics.TR_.pdf
- <https://www.statista.com/statistics/308330/number-stream-users/#:~:text=Steam%20had%20approximately%20120%20million,monthly%20active%20users%20in%202019.>
- <https://www.statista.com/topics/4282/steam/#dossierKeyfigures>
- <https://www.statista.com/statistics/823933/steam-retention/>
- <https://www.statista.com/statistics/308330/number-stream-users/#:~:text=Steam%20had%20approximately%20120%20million,monthly%20active%20users%20in%202019.https://www.statista.com/statistics/308330/number-stream-users/#:~:text=Steam%20had%20approximately%20120%20million,monthly%20active%20users%20in%202019.>