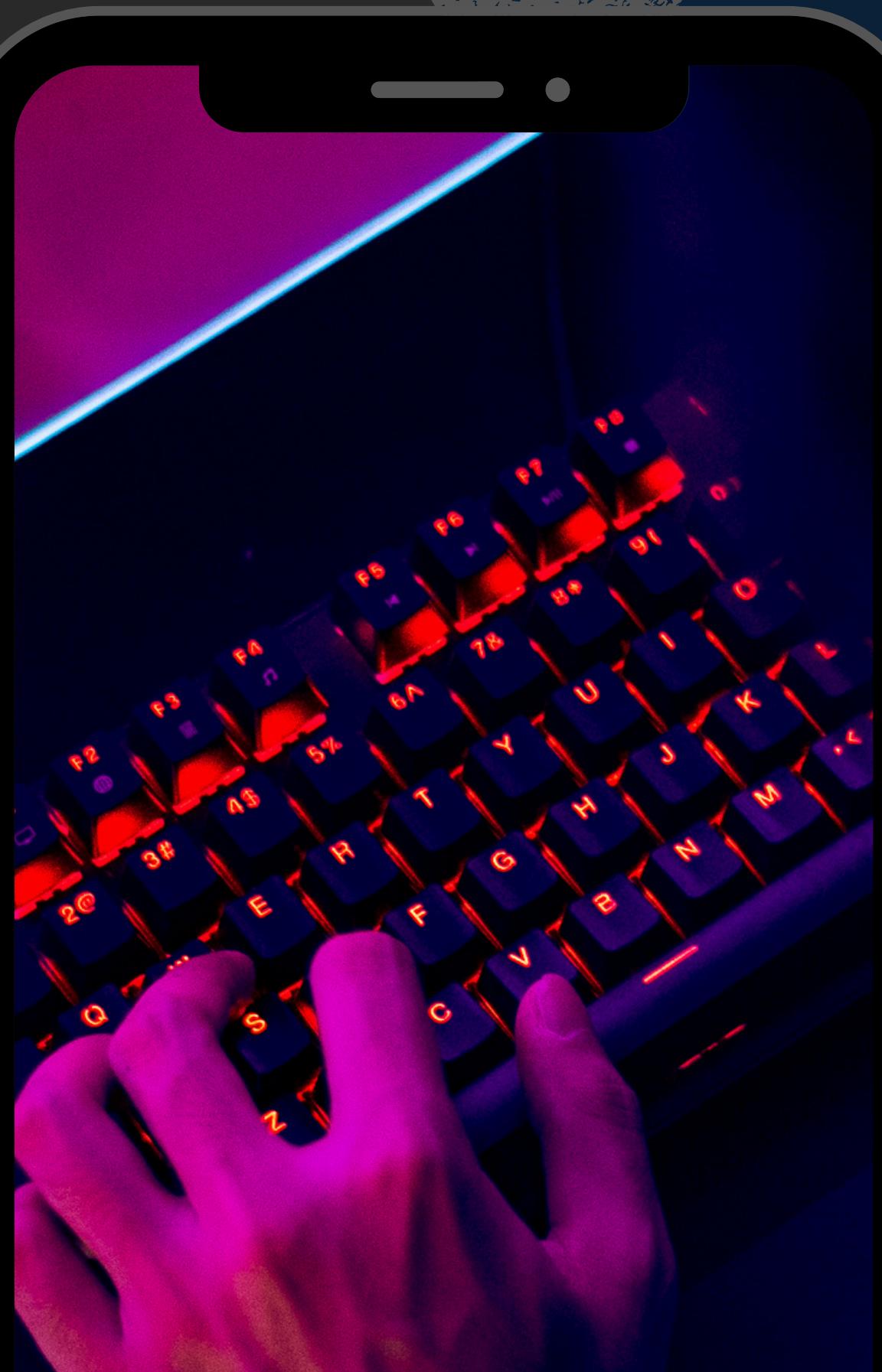


Do Reddit and Twitch Trends Reflect Steam Player Counts? A Data Science Investigation

Mini Project 3
IOD - 02- 27 Cohort

Created by
Ryo Suda



Business Problem

Can social engagement signals—like Reddit sentiment and Twitch streaming activity—help predict or explain player counts for top Steam games? Understanding these relationships is valuable for game publishers and marketers seeking to grow player engagement



MOST PLAYED

Top 100 played games By Current Players

RANK

RANK	GAME	PRICE	CURRENT PLAYERS	PEAK TODAY
1	Counter-Strike 2 (hidden by your preferences)	Free To Play	673,410	1,305,403
2	Dota 2 (hidden by your preferences)	Free To Play	321,559	658,458
3	PUBG: BATTLEGROUNDS (hidden by your preferences)	Free To Play	250,049	715,892
4	Bongo Cat	Free To Play	134,545	163,477
5	Delta Force	Free To Play	102,441	197,482
6	Marvel Rivals	Free To Play	92,589	139,338
7	Wallpaper ENGINE			

Browse

Games

IRL

Music & DJs

Creative

Esports

Categories Live Channels



Just Chatting

267K viewers

IRL



League of Legends

49.7K viewers

RPG Strategy



Marvel Rivals

23.1K viewers

Shooter Action



Old School RuneScape

7.8K viewers

FPS Adventure Game



Counter-Strike

18.7K viewers

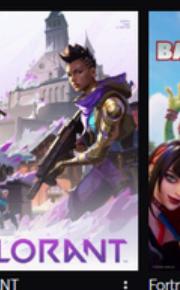
FPS Shooter Action



IRL

94.1K viewers

IRL



VALORANT

37.2K viewers

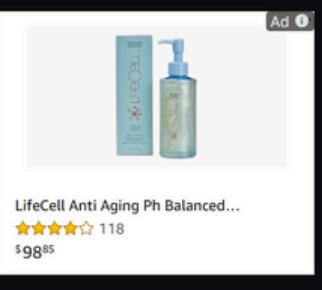
FPS Shooter Action



Fortnite

28.8K viewers

Shooter RPG



LifeCell Anti Aging Ph Balanced...

4.5★ 118

\$98.85



PEAK

7.7K viewers

Platformer



Warframe

1.9K viewers

Shooter RPG



Minecraft

12.4K viewers

Simulation



Destiny 2

34.4K viewers

FPS Shooter RPG



World of Warcraft

9.2K viewers

RPG Adventure Game



Grand Theft Auto V

31.9K viewers

FPS Shooter RPG



Escape from Tarkov

16.2K viewers

FPS Shooter RPG



Donkey Kong Bananza

6.6K viewers

Platformer



Dead by Daylight

19.6K viewers

Strategy Action



Umanusume: Pretty

4.7K viewers

RPG Simulation



Call of Duty: Warzone

26K viewers

FPS Shooter



Art



ROBLOX



Rainbow 6



Haute Shot



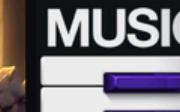
Teamfight Tactics



Clash Royale



Overwatch 2



Expedition



Music

Search r/gaming × Search in r/gaming

Community highlights

Simple Questions Sunday! 3 votes • 21 comments

u/Extasio • 3 hr. ago Australian anti-porn group claims responsibility for Steam's new censorship rules in victory against 'porn sick brain rotted pedo gamer fetishists', and things only get weirder from there

www.pcgamer.com/gaming-industry/australian-anti-porn-group-claims-respons...

ncsa222 • 11 hr. ago a Fantasy Japanese City in Minecraft



Project Objective

Analyze whether Reddit comment scores and Twitch stream counts correlate with or predict Steam player activity, and determine which metrics are most useful for modeling and segmenting game popularity.

Questions Addressed:

Are Reddit and Twitch metrics associated with Steam player counts?
Can these metrics be used for predictive modeling or game segmentation?

Disclaimer

This project originally aimed to analyze a large dataset of over 290,000 unique games acquired via the Steam API. During exploratory analysis, we encountered issues such as the inclusion of mods (fan-published content) and DLCs (game extensions), as well as data restrictions for adult-rated games and inconsistencies in game records. As a result, we decided not to proceed with this dataset.

We also explored using a Steam dataset from Kaggle, but faced difficulties in gathering corresponding Twitch data—many titles were too old, leading to missing or zero values for Twitch engagement. Attempts to supplement this with Reddit data were unsuccessful due to API limitations in retrieving relevant posts for these games.

Given these challenges, we narrowed our scope to a focused dataset of the top 20 games on Steam. For these games, we were able to successfully collect Twitch live user counts and Reddit top comments and likes, which formed the basis of our analysis. Despite its restricted statistical power, we were able to demonstrate the key objectives of the project.

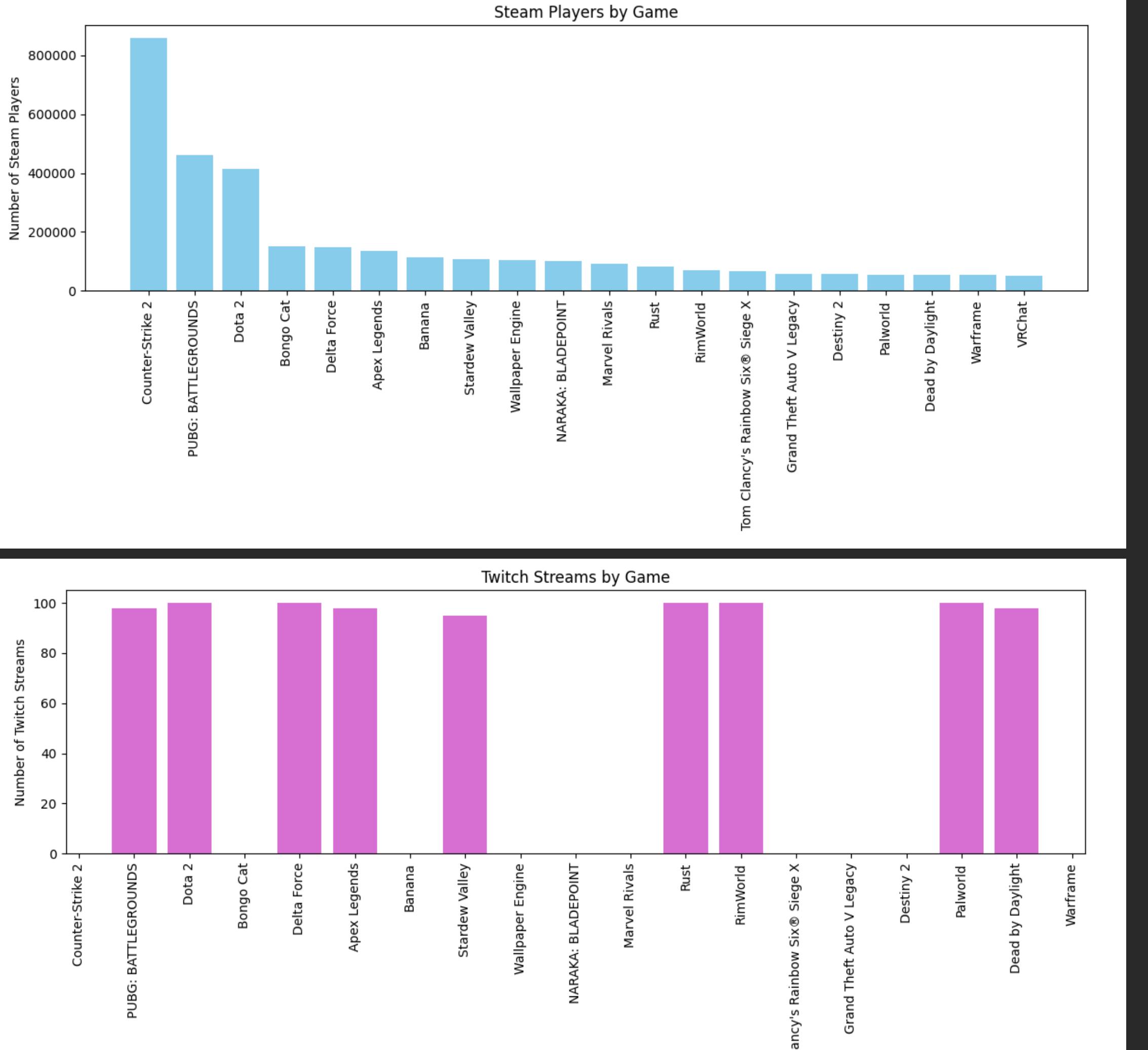
Data Collection

Gathered the current top 20 games on Steam via API.
Retrieved Twitch live stream counts for these games.
Collected the top Reddit comment and its score for each game.

- **Game** - The title or name of the game being analyzed.
- **AppID** - Unique identifier for each game on the Steam platform
- **Steam Players** - The current number of active players for the game on Steam
- **Twitch Streams** -The count of live Twitch streams currently broadcasting the game
- **Reddit Mentions** -The number of times the game is mentioned in relevant Reddit posts
- **Reddit Comment Score** -The score (likes/upvotes) of the top Reddit comment for each game
- **Top Reddit Comment** -The content of the highest-voted Reddit comment related to the game



Top 20 Games Player Count and Twitch Mentions



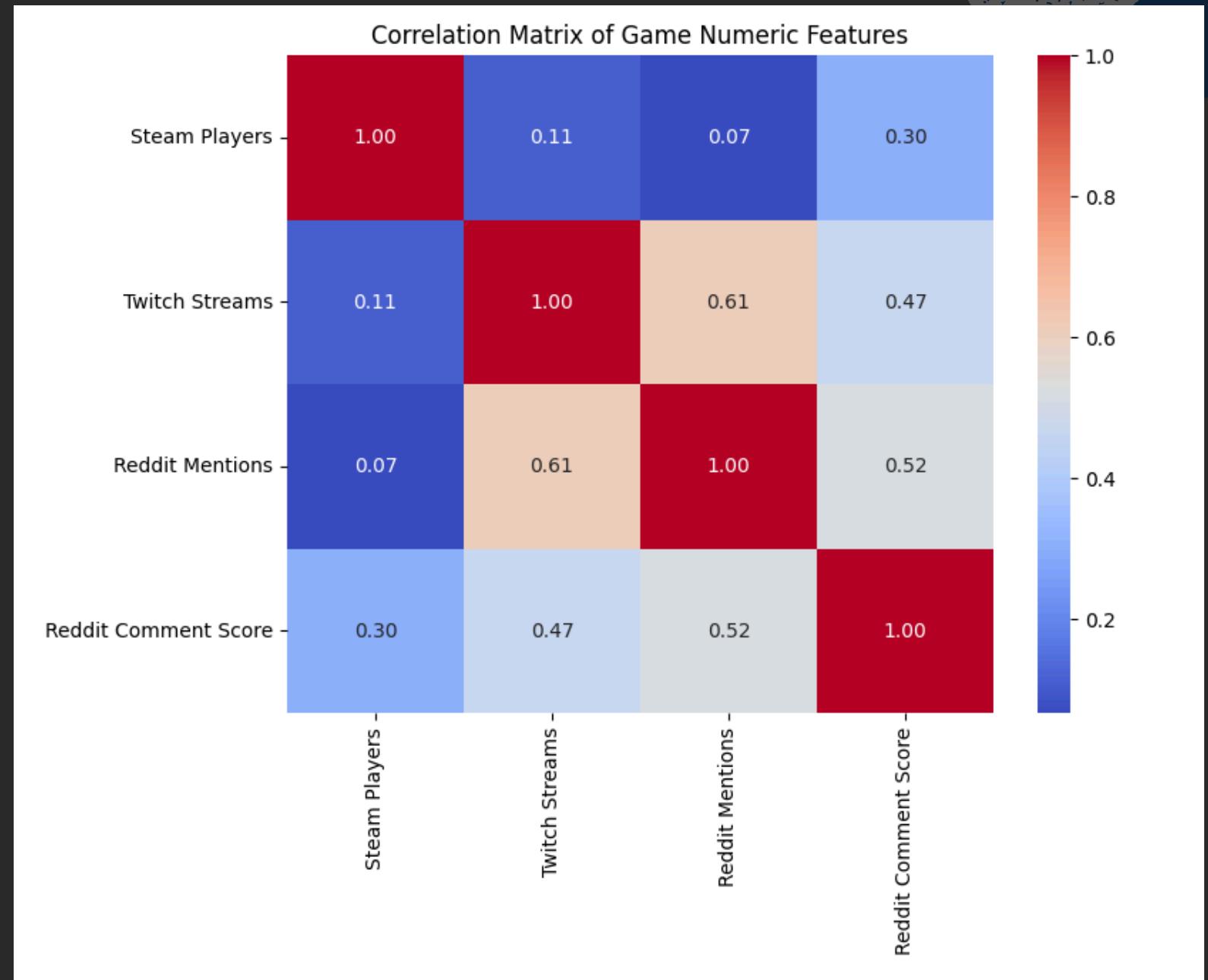
Correlation Analysis

We computed the correlation matrix for the main numeric features (Steam Players, Twitch Streams, Reddit Mentions, Reddit Comment Score). The most significant findings:

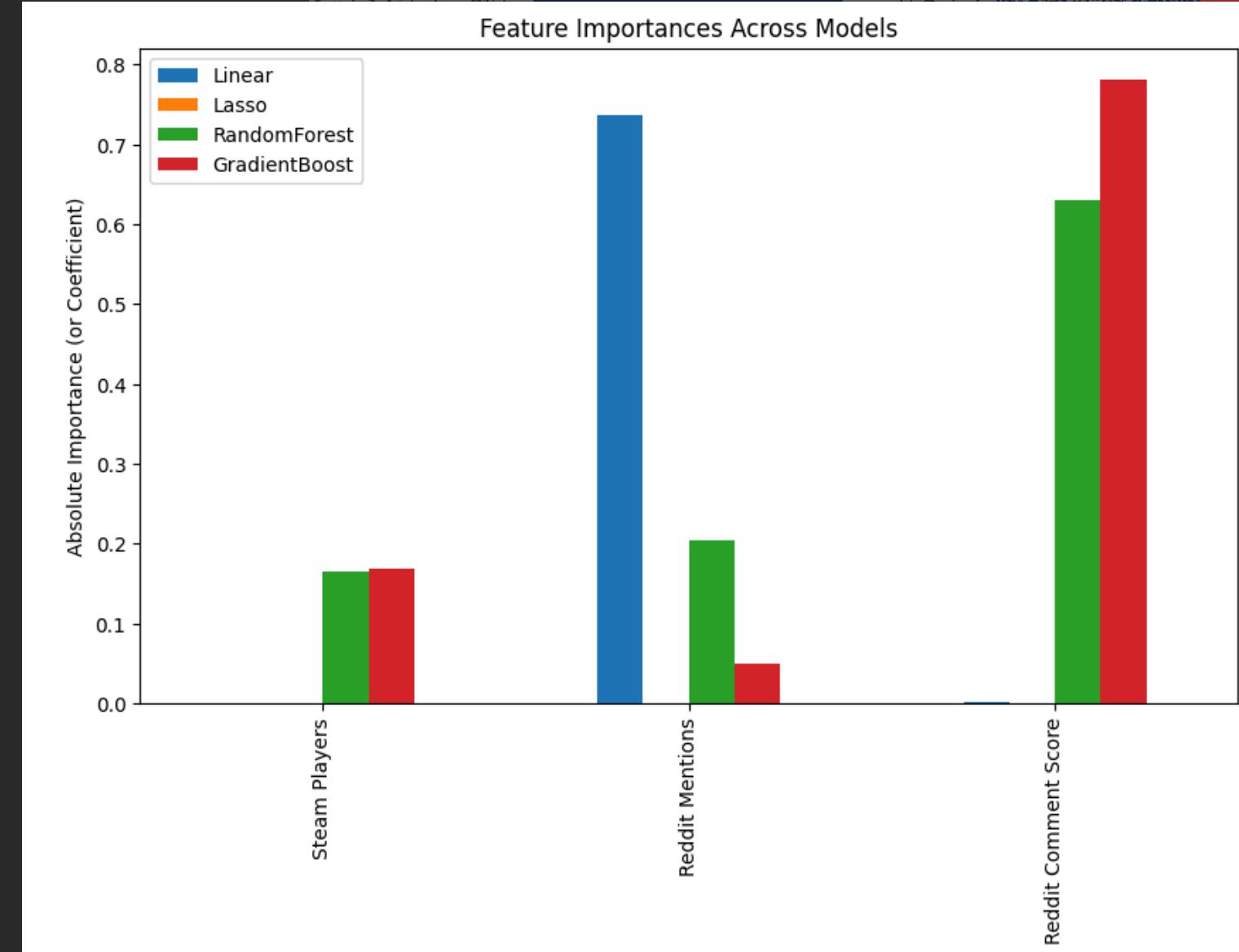
Twitch Streams and Reddit Mentions: Correlation = 0.61
(This suggests that games popular on Reddit are also popular on Twitch)

Reddit Mentions and Reddit Comment Score: Correlation = 0.52
(Games with more Reddit activity tend to have more highly scored comments)

Steam Players is not strongly correlated with Twitch or Reddit metrics, indicating that in this sample, top Steam player counts do not necessarily translate to social or streaming buzz.

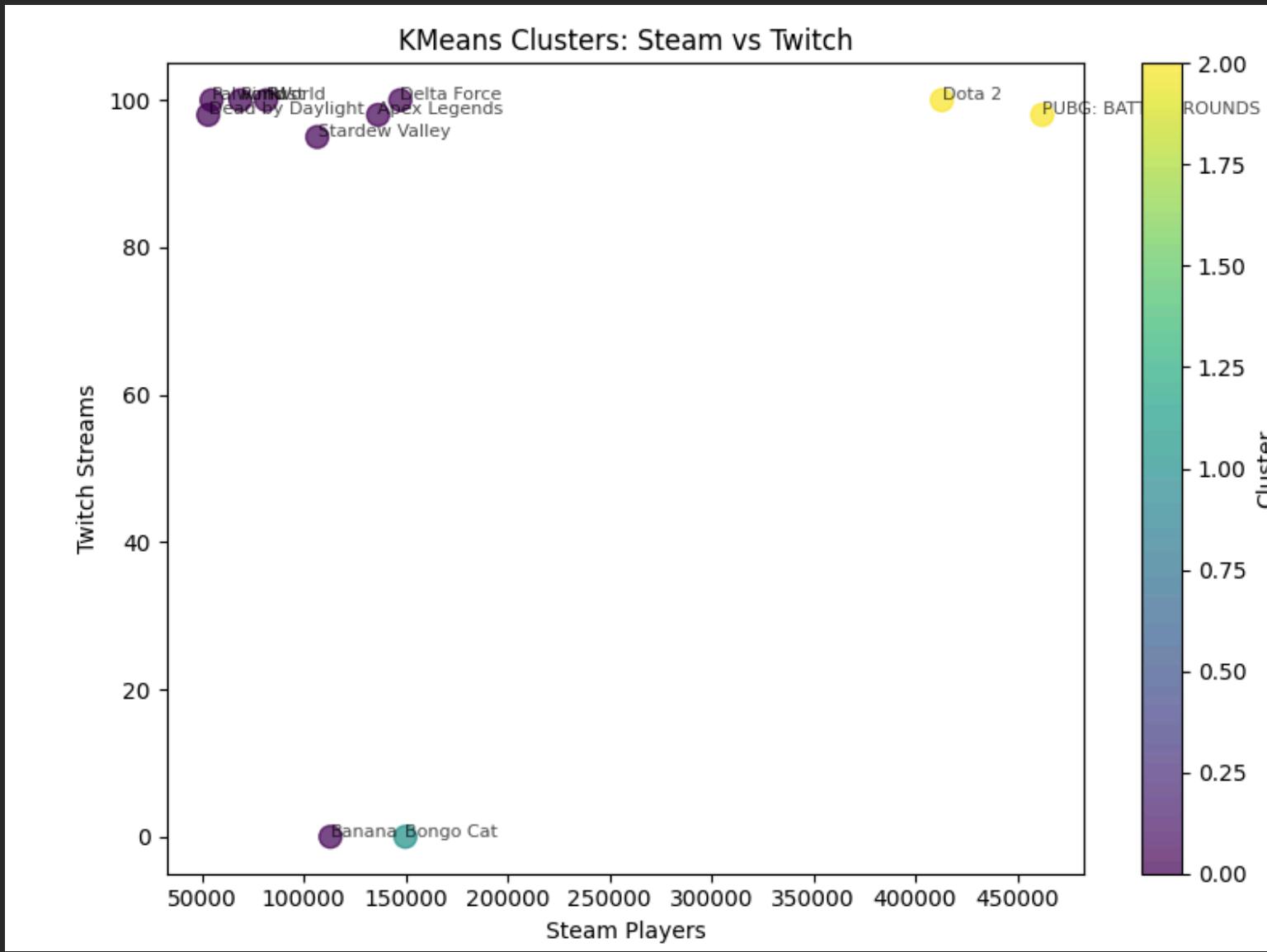


Modeling & Feature Importance



Model/Method	Feature	Importance / Coefficient	Notes/Insights
Random Forest / Gradient Boosting	Reddit Comment Score	Highest	Most important predictor for Twitch Streams.
	Reddit Mentions	Moderate	Next most important predictor.
	Steam Players	Lowest	Least predictive feature.
Linear Regression	Reddit Mentions	Highest	Most important predictor (largest coefficient).
	Reddit Comment Score	Very Low	Weak coefficient.
	Steam Players	Very Low	Weak coefficient.
Lasso Regression	All features	0	All coefficients set to zero, indicating over-regularization (likely due to sample size/redundancy).

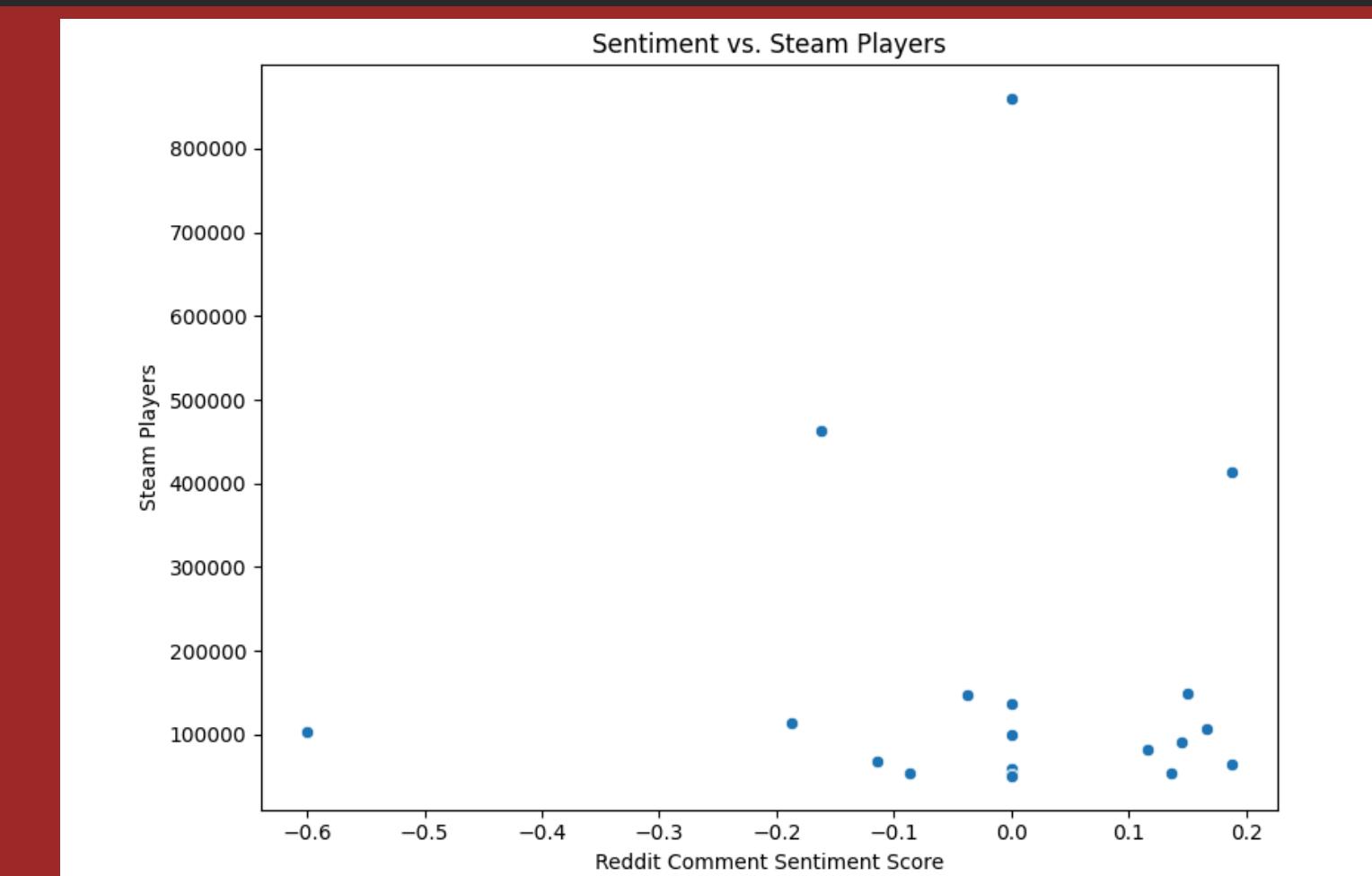
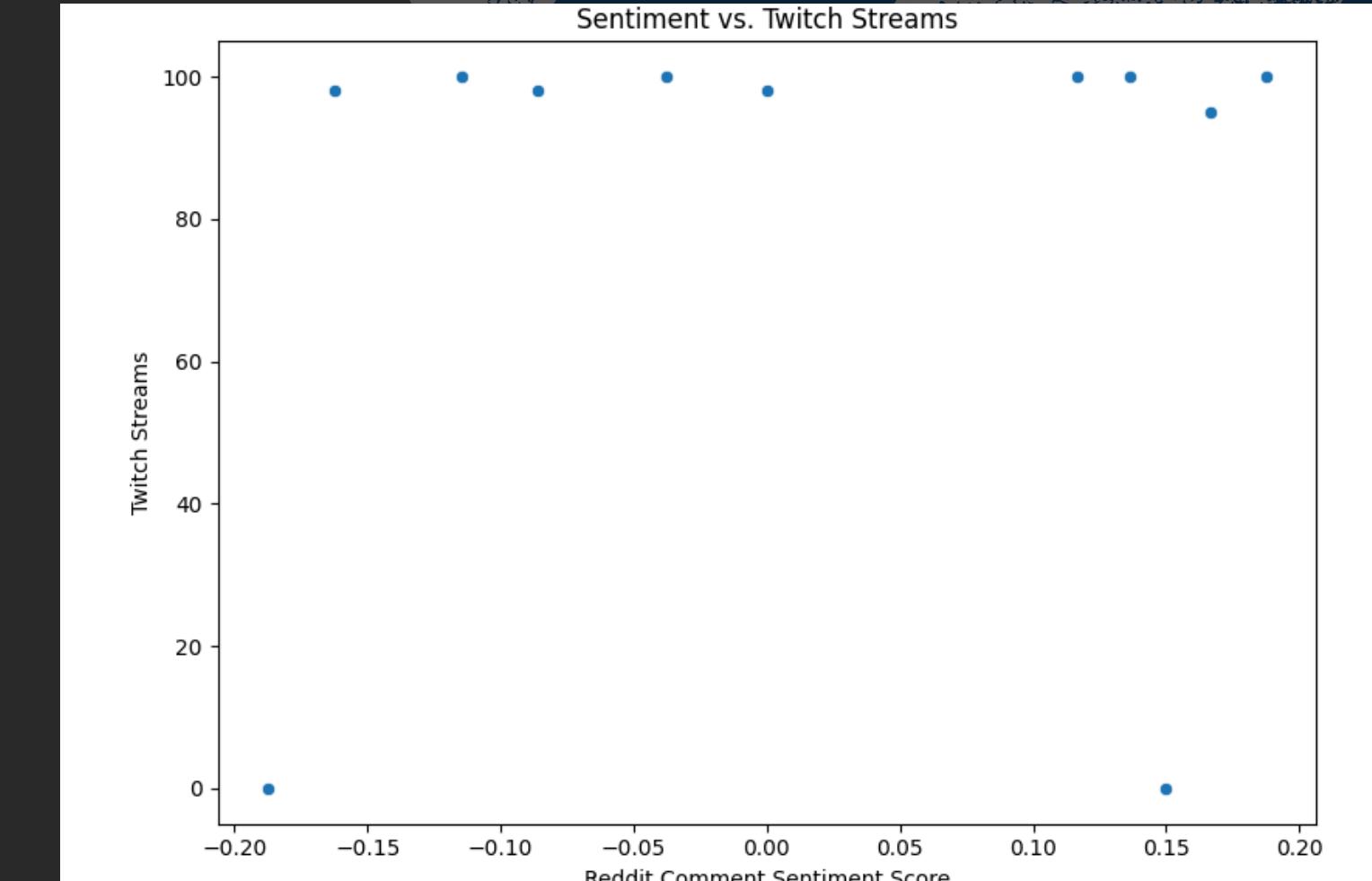
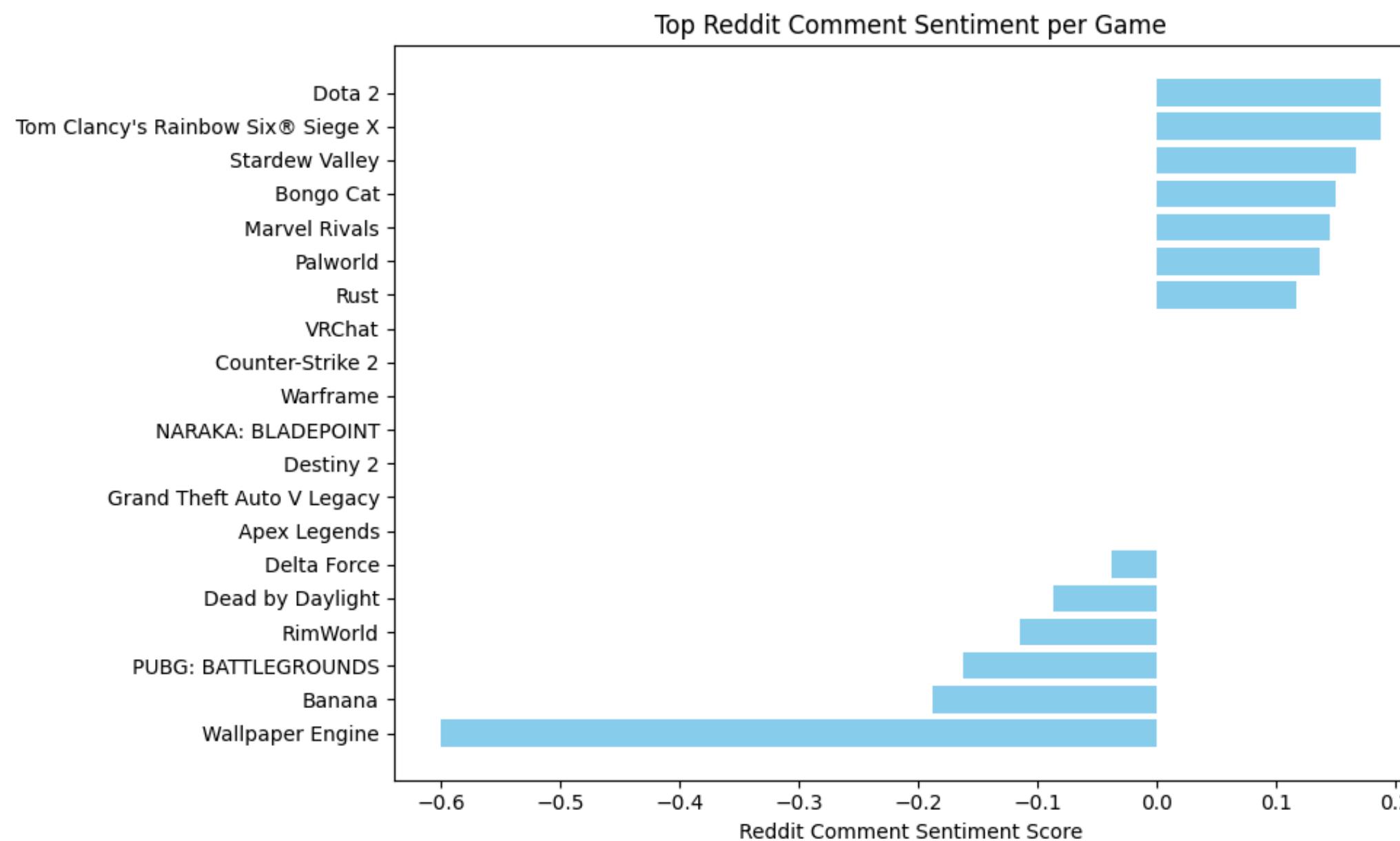
Clustering



Cluster	Steam Players	Twitch Streams	Reddit Mentions	Reddit Comment Score	Typical Profile
0	95,301	86.4	100	5,656	Mid-popularity, high Reddit activity, above-average Twitch
1	149,916	0	3	6	Single standout game (likely Bongo Cat), high Steam, no Twitch/Reddit
2	437,692	99	84.5	5,080	Top-tier games, very high Steam & Twitch, strong Reddit

Sentiment Analysis

We used Vadar and texblot which gave scores on the sentiment analysis saw the following games with positive sentiment. It also seems to suggest sentiment for high current players are likely positive. But it does not interpret the same for Twitch Streams



In Conclusion

1. **Community-driven metrics (Reddit, Twitch) can be more indicative of streaming engagement than raw player counts alone.**
2. **We need computing power and API access to allow larger data to be analyzed**

Lessons and more things to explore

1. **There is limitation API data, study ways around such as third party providers.**
2. **Gather more Reddit Comments to strengthen the Sentiment Analysis and Correlation to Number Features**



**Thank you for
listening**