

Steam Games Review Analysis with Hadoop MapReduce

Isai Tinoco Gutierrez, Russell Ferrall

Project Motivation & Goals

- Motivation: Why analyze Steam Reviews dataset? (large, diverse, interesting for analysis).
- Goal: Identify highest/lowest rated games & games with most reviews.
- Tool: Hadoop MapReduce (distributed processing).

Dataset Description

- Source: Kaggle Steam Games Reviews (2024).
- <https://www.kaggle.com/datasets/artermiloff/steam-games-reviews-2024>
- Size: 128M reviews, 80k games, 30M+ users, >5 GB.
- Attributes used: (appid, language, voted_up, early_access)

Data Cleaning

- Filters applied in MapReduce Mapper:
 - English reviews only.
 - Exclude early access reviews.
 - Remove corrupted/malformed rows.

System Setup

- Hadoop 3.4.1 on pseudo-distributed mode.
- Input data stored in HDFS
- JAR built with Maven (pom.xml)
- Diagram: **HDFS Input -> Mapper -> Combiner -> Reducer -> HDFS Output.**

MapReduce Implementation

- Mapper: emits (appid, (positive, total)).
- Combiner: aggregate partial (sum, count)
- Reducer: compute average = $\text{sum} / \text{count}$ per game

Hadoop Execution

- Input in HDFS: /user/\$USER/steam/SteamReviews2024
- Output in HDFS: /user/\$USER/steam/out_appstats_
- Example commands: `hadoop jar target/HadoopLab-1.0-SNAPSHOT.jar`
- Cluster mode: pseudo-distributed

Demo

The screenshot shows a terminal window titled "Terminal" with a menu bar (Shell, Edit, View, Window, Help) and a status bar (100%, Mon Aug 25 9:03 PM). The terminal content displays a web application interface for "HadoopLab". The interface is dark-themed and shows the following details:

- Application: **application_1756158588323_0006**
- Logged in as: dr.who
- Application Description: (faded text)
- Application Status: (faded text)
- Application Resource: (faded text)
- Application Preempted: (faded text)
- Application Killed: (faded text)
- Application Resource Allocation: (faded text)
- Application Preempted Resource Allocation: (faded text)
- Table with 4 columns: Node, Logs, Nodes blacklisted by the app, Nodes blacklisted by the system.

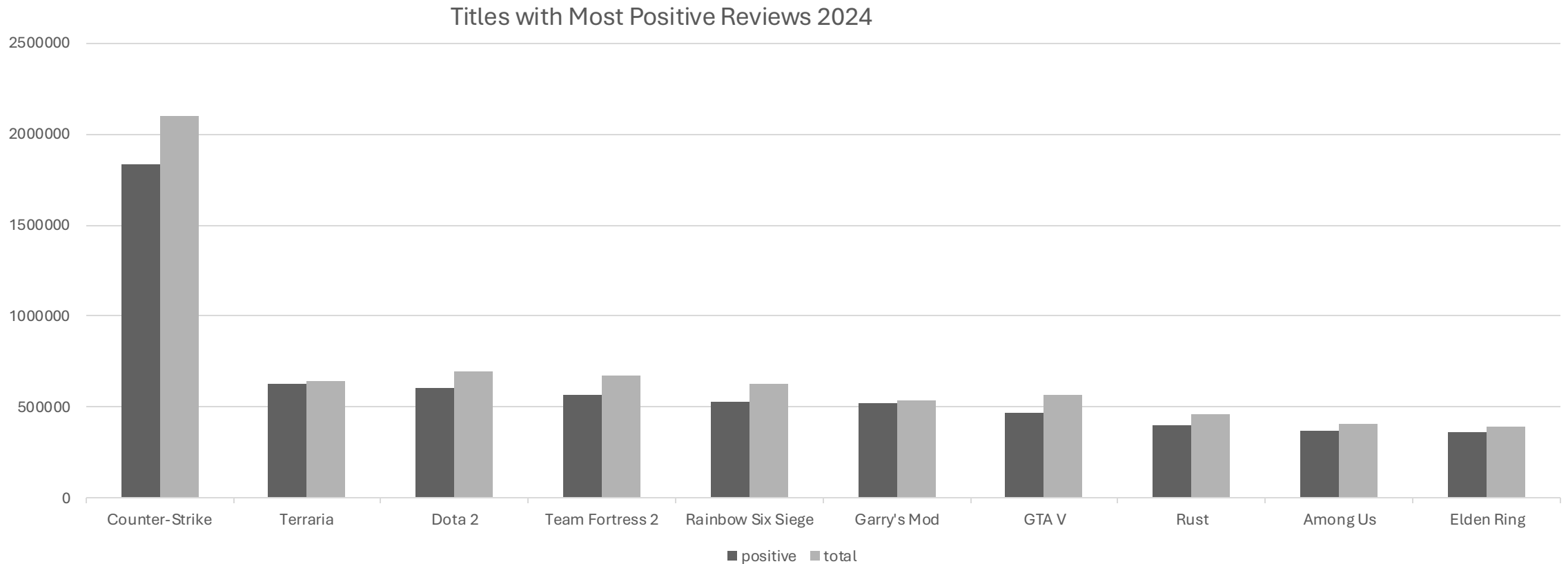
Node	Logs	Nodes blacklisted by the app	Nodes blacklisted by the system
http://mac:8042	Logs	0	0

Page navigation: First, Previous, 1, Next, Last

Results: Most Positive Reviews

- Chart: "Titles with Most Positive Reviews 2024".
- Key findings: *Counter-Strike* leads (~ 1.9M positives).
- Other strong games: *Terraria*, *Dota 2*, *Rainbow Six Siege*.

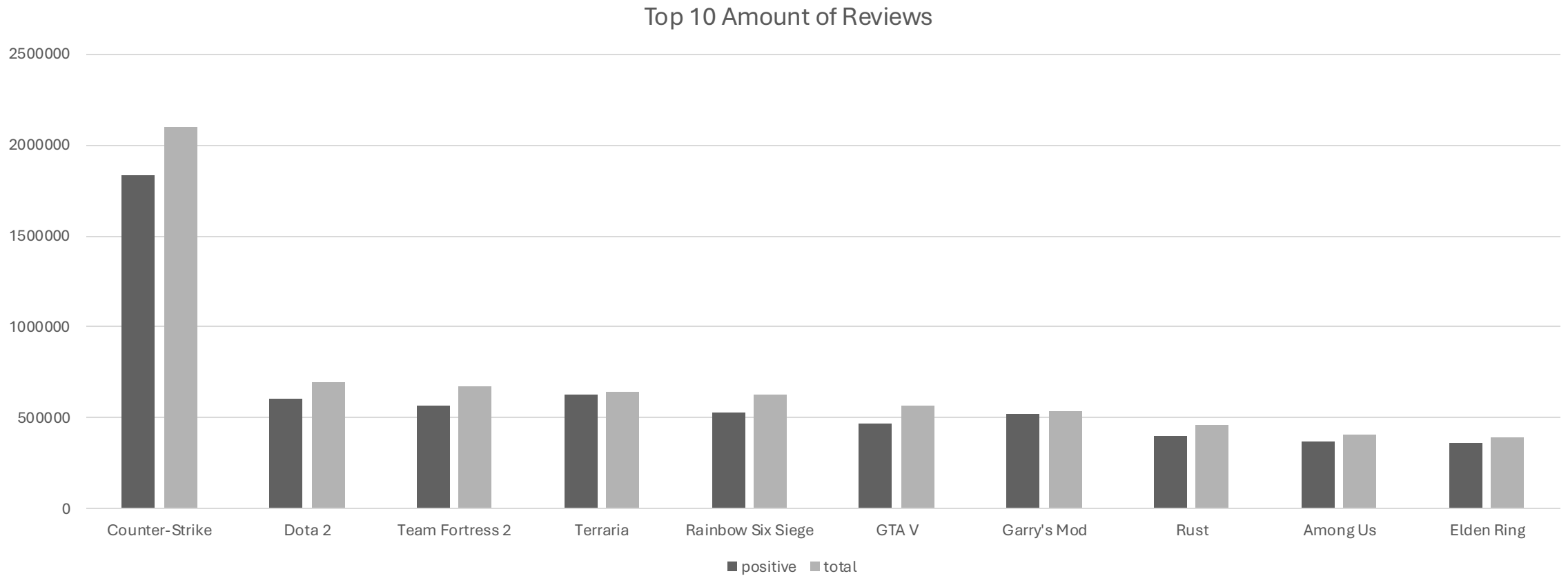
Most Amount of Positive Reviews



Results: Most Reviewed Games

- Chart: "Top 10 Amount of Reviews".
- *Counter-Strike & Dota 2* dominate review counts.
- *Among Us & Elden Ring* appear with high review activity.

Most Amount of Reviews

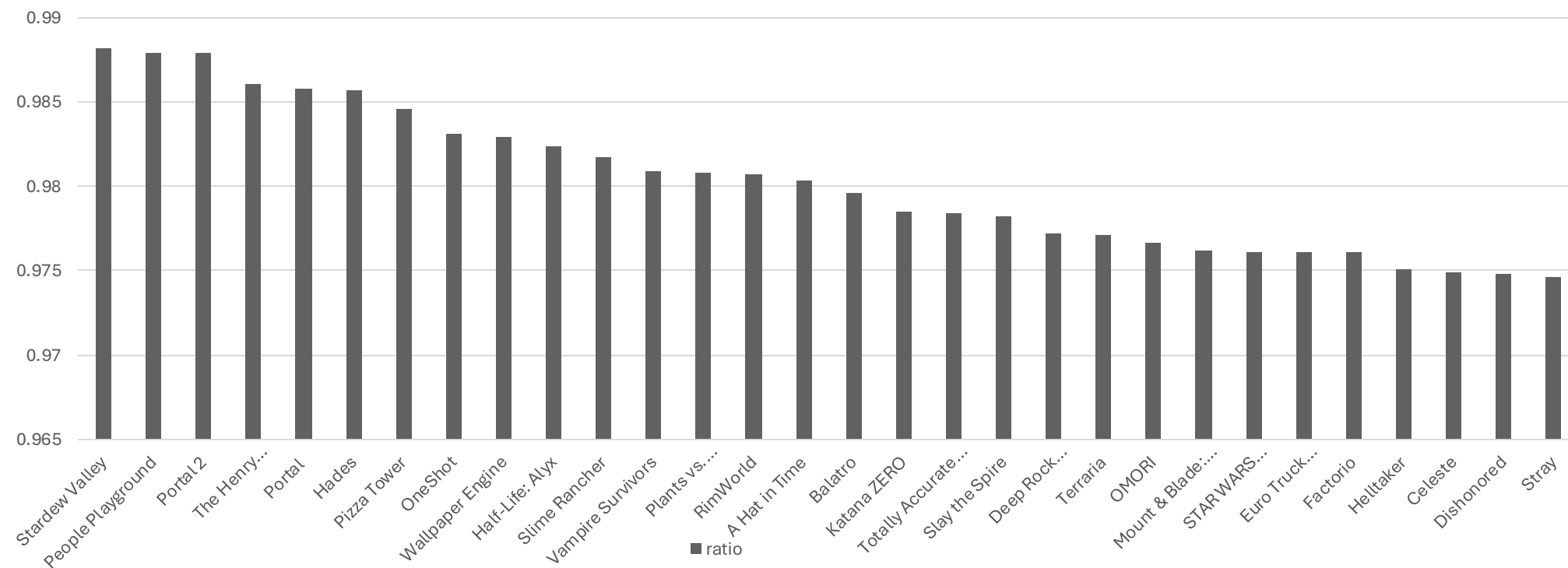


Insights

- Community-driven games (e.g., *Counter-Strike*, *Dota 2*) maintain massive engagement.
- Newer titles (*Elden Ring*, *Among Us*) competitive but smaller.
- Ratio of positive to total reviews highlights reputation differences.

Highest Review Ratio

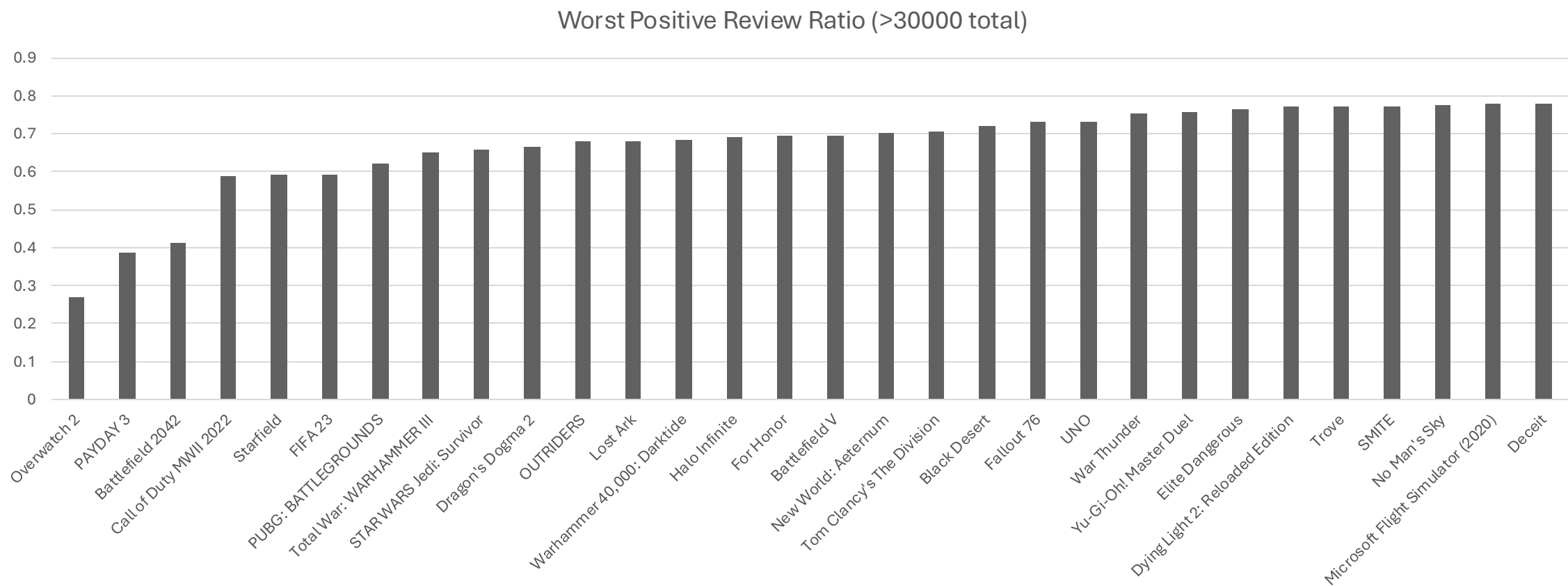
Games with Highest Positive Reviews (> 30000 total reviews)



Highest Review Ratio (2024)

- Many games show extremely high positive review ratios (97.5%+).
- *Stardew Valley* and *People Playground* lead with nearly 99% positive reviews.
- Indicates that popular indie and community-driven games often receive overwhelmingly positive feedback.

Worst Review Ratio



Worst Review Ratio (2024)

- A few games had much lower positivity, with *Overwatch 2* falling under 25%.
- Only 3 games dropped below 50% positive reviews overall (*Overwatch 2*, *PAYDAY 3*, *Battlefield 2042*), meaning most retain strong player approval.
- Suggests that community backlash or poor updates can drastically affect ratings despite high visibility.

Key Takeaways

- In 2024, most games received overwhelmingly positive reviews, Top 30 games barely dipped under 97.5% positivity.
- For games with more than 30,000 reviews, only 3 dropped below 50% positivity.
- This contrast highlights how **most games trend positivity**, but a few major titles experience strong community dissatisfaction.

Conclusion

- **Most Popular Games:**
- *Counter-Strike*, *Terraria*, and *Dota 2* dominate in both total reviews and positive feedback.
- **Positivity Trends:**
- Majority of games in 2024 maintain **97.5%+ positive reviews**, showing overall satisfaction.
- **Outliers:**
- Only a handful of games (*Overwatch 2*, *Battlefield 2042*) fall below **50% positivity**, highlighting cases of strong backlash.
- **Takeaway:**
- Steam reviews reflect **both loyalty and frustration** – successful games build trust, while poorly received updates erode it quickly.