

# 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>

recommendationid	language	timestamp_created	timestamp_updated	voted_up	votes_up	votes_funny	weighted_vote_score	comment_count	steam_purchase	received_for_free	written_during_early_access	hidden_in_steam_china
36638296	brazilian	1511149221	1695980563	0	0	0	0.5454545617103580	0	1	0		1

- Size: 128M reviews, 80k games, 30M+ users, >14 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.**

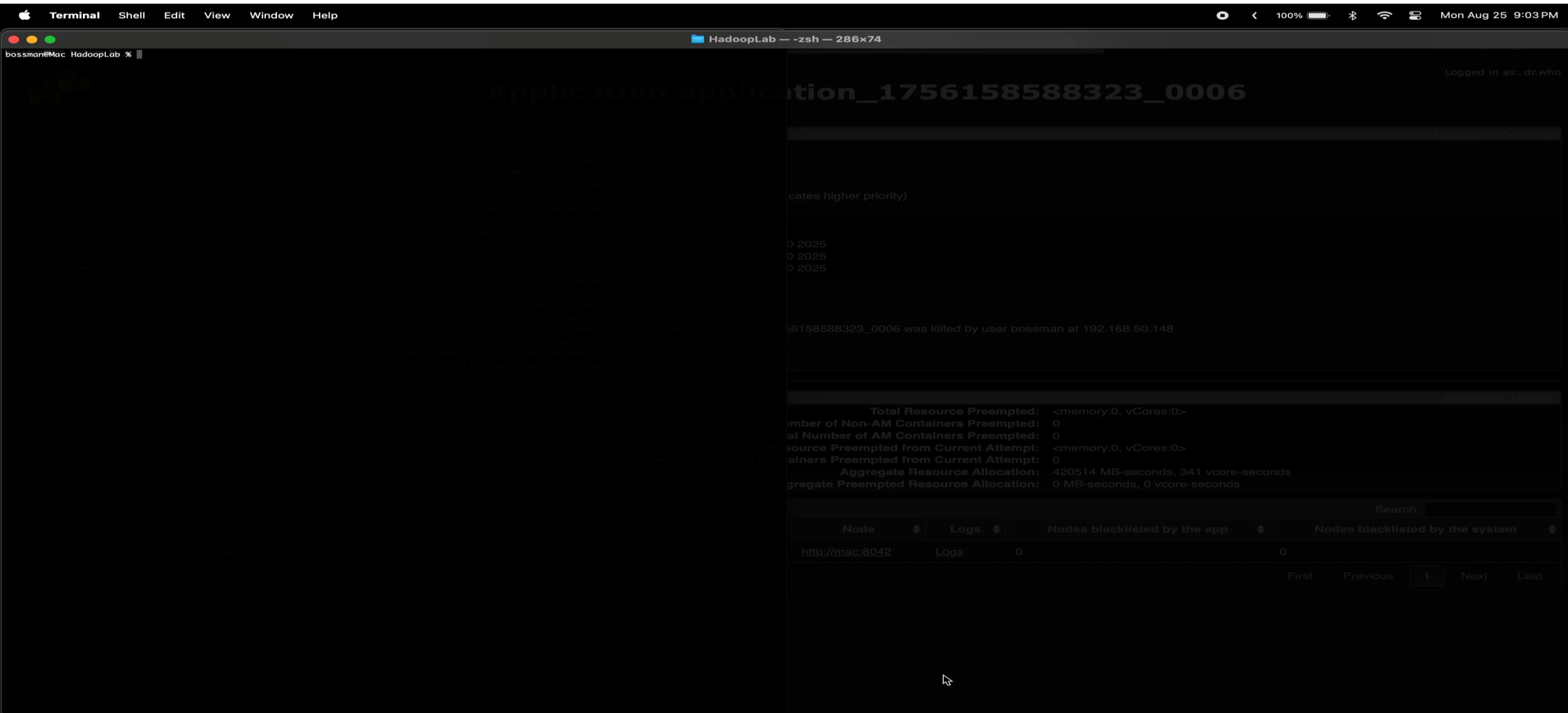
# MapReduce Implementation

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

# Hadoop Execution

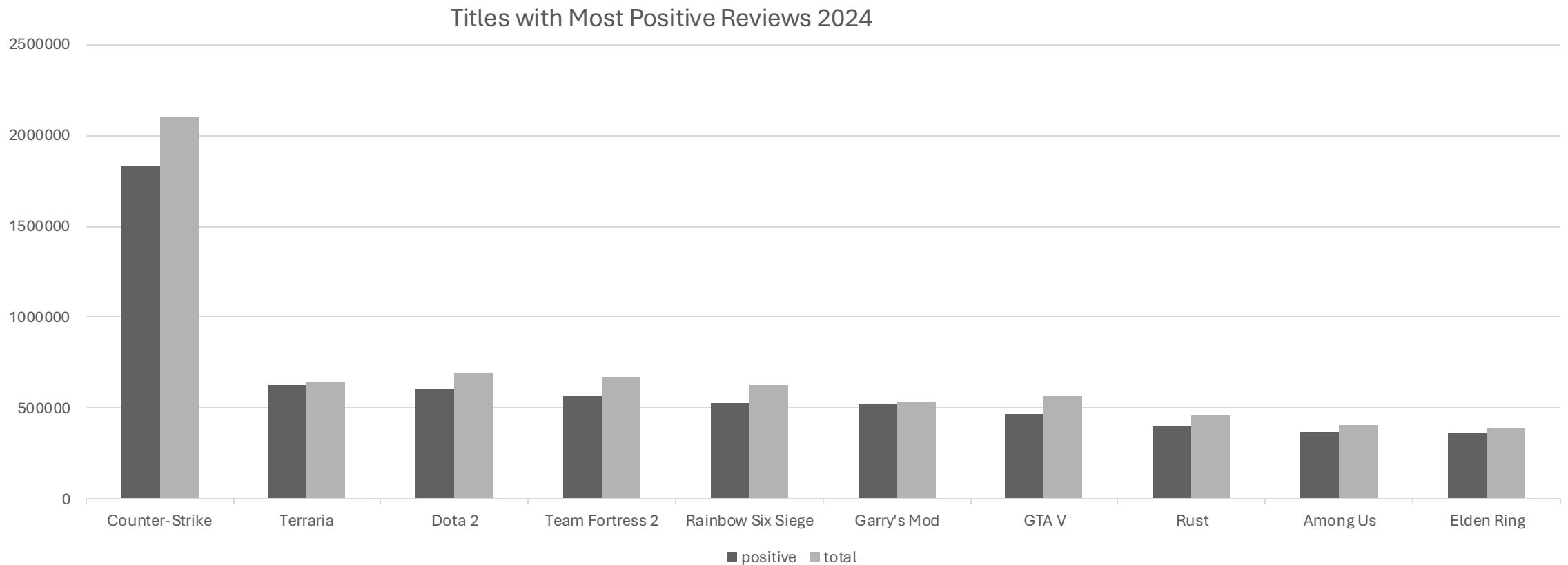
- Input in HDFS: `/user/$USER/steam/SteamReviews2024`
- Output in HDFS: `/user/$USER/steam/out_appstats_`
- Example commands: `hadoop jar target/HadoopLab-1.0-SNAPSHOT.jar`
- Cluster mode: pseudo-distributed

# Demo





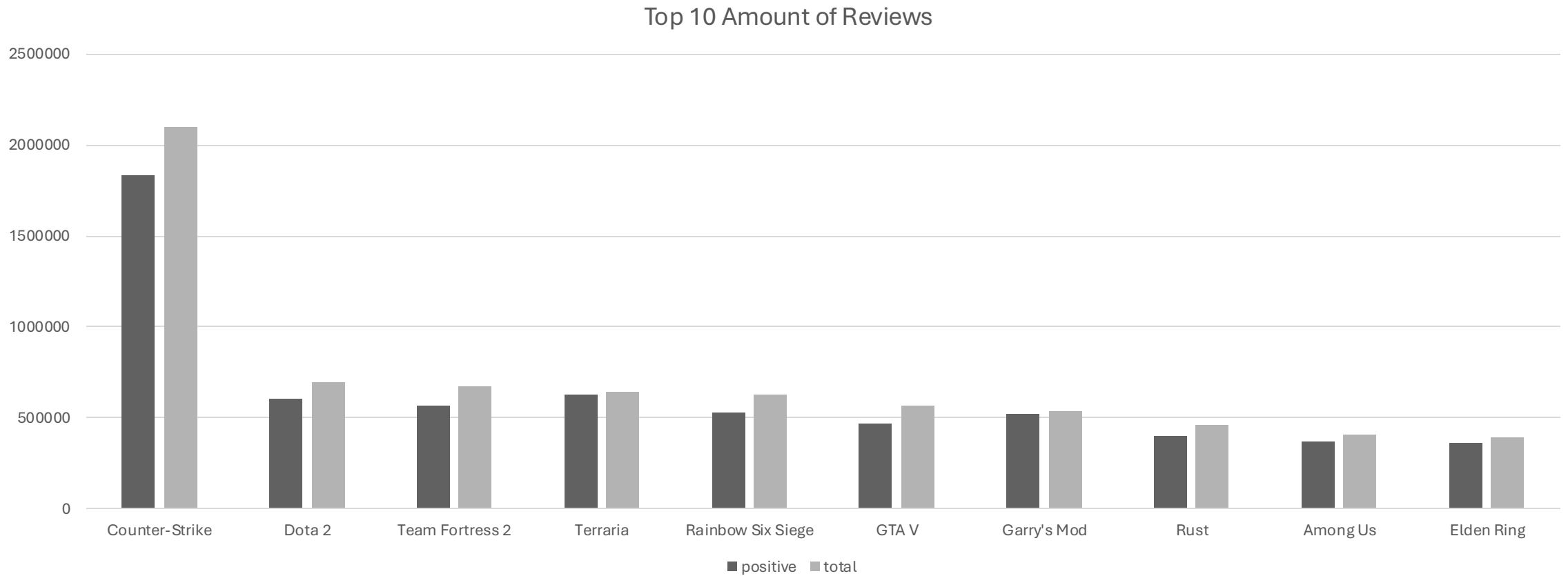
# Most Amount of Positive Reviews



# 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 Reviews



# Results: Most Reviewed Games

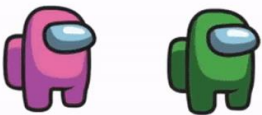
---

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



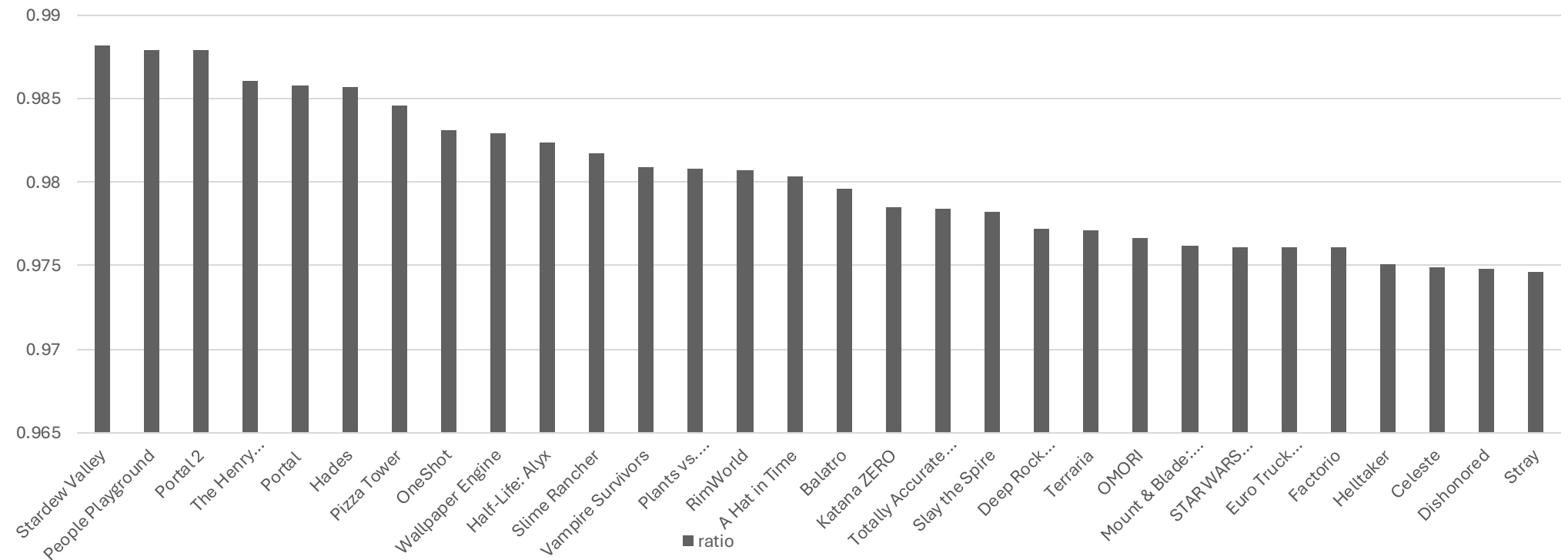
# Insights

- 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.



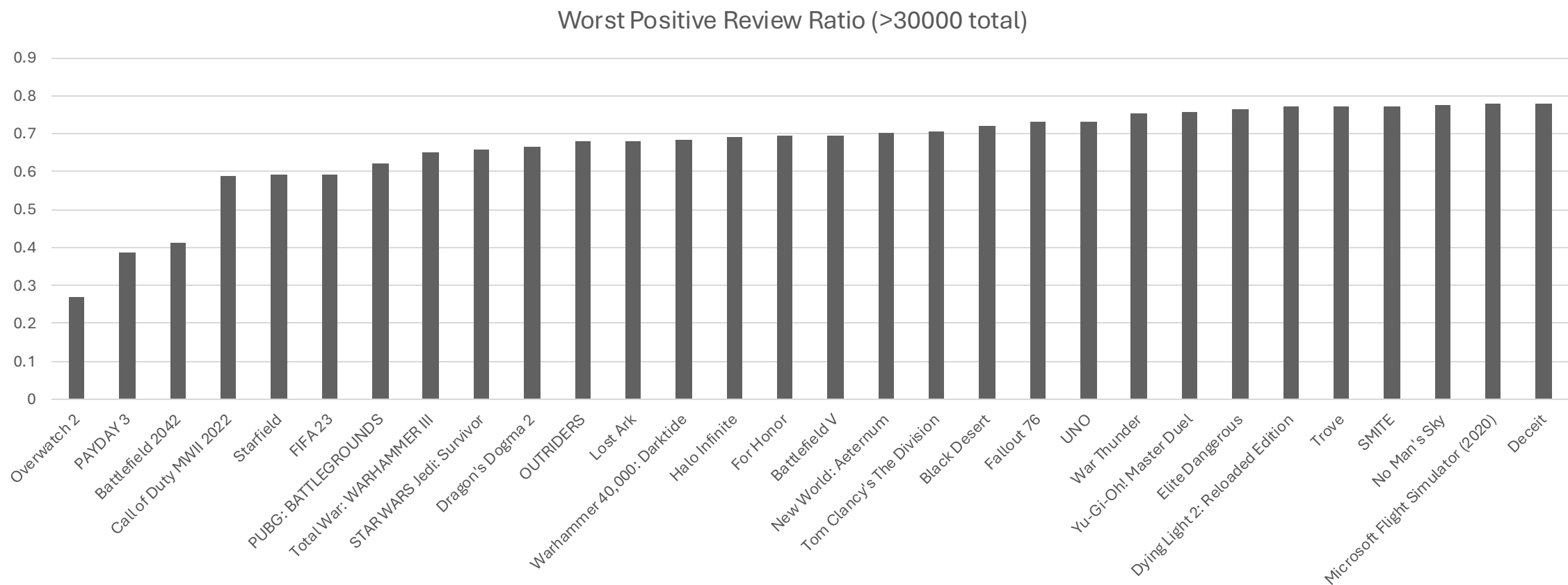
Dig, fight, explore, build! Nothing is impossible in this action-packed adventure game. Four Pack also available!

RECENT REVIEWS: **Overwhelmingly Positive** (21,883)  
ALL REVIEWS: **Overwhelmingly Positive** (250,222)

RELEASE DATE: May 16, 2011

DEVELOPER: Re-Logic  
PUBLISHER: Re-Logic

# 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.