

# Video Game Recommendations

As a long-time video game fan, I want to understand the kinds of games that sell well.  
This could allow us to **predict successful games** (to improve profits for a company).  
We can also use this information to **recommend games to users**.



## Goals:

1. Understand the factors predictive of video game success.
2. Build a video game recommender based on this data.

## Code:

[https://github.com/aedanli/1\\_GamesRecommendation](https://github.com/aedanli/1_GamesRecommendation)

**Aedan Yue Li, Ph.D.**

# Exploring a Steam Game Dataset

I analyzed a recent dataset of Steam Games with 76,000+ titles scraped on Aug 8, 2023.

**Dataset:** <https://www.kaggle.com/datasets/fronkongames/steam-games-dataset>



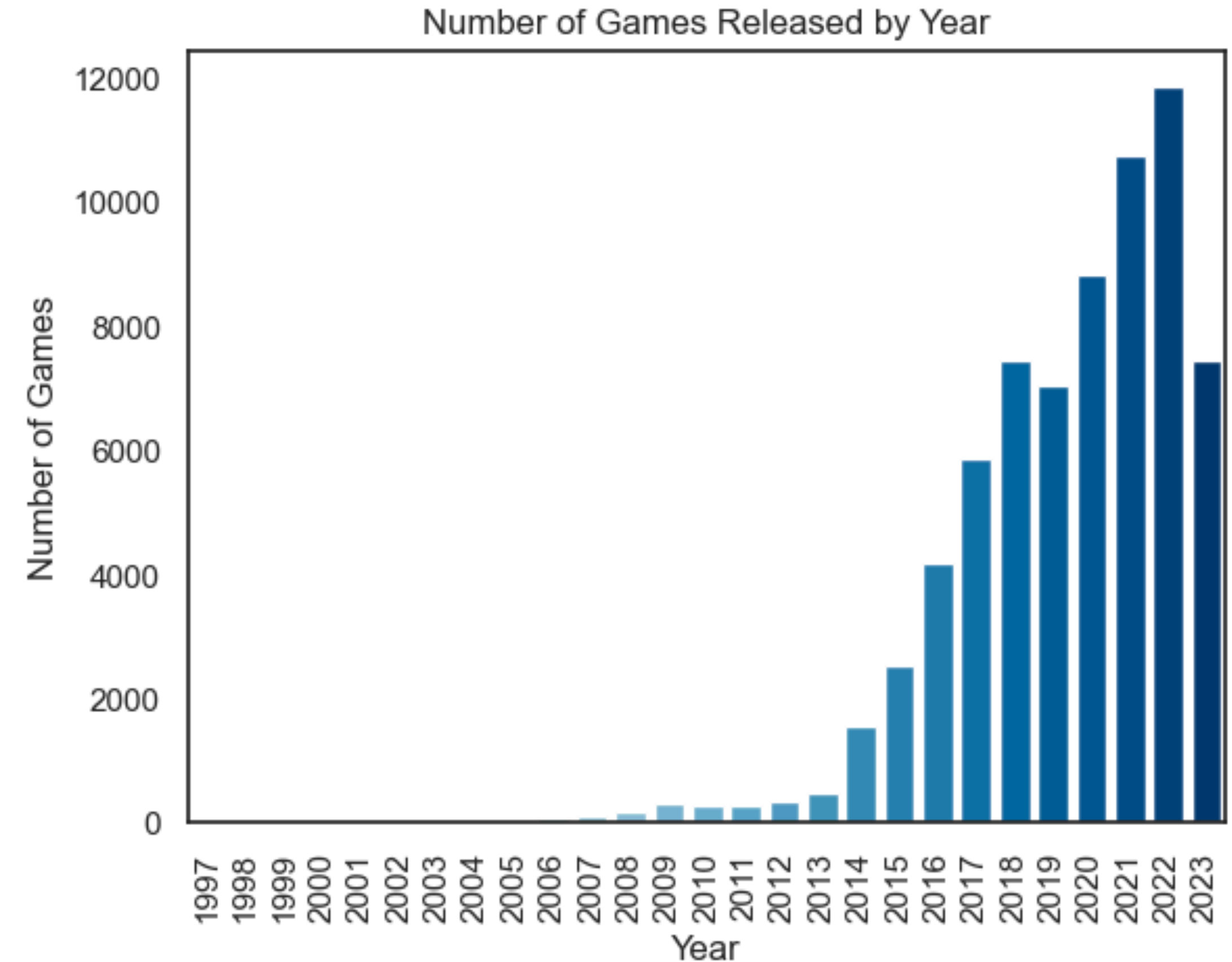
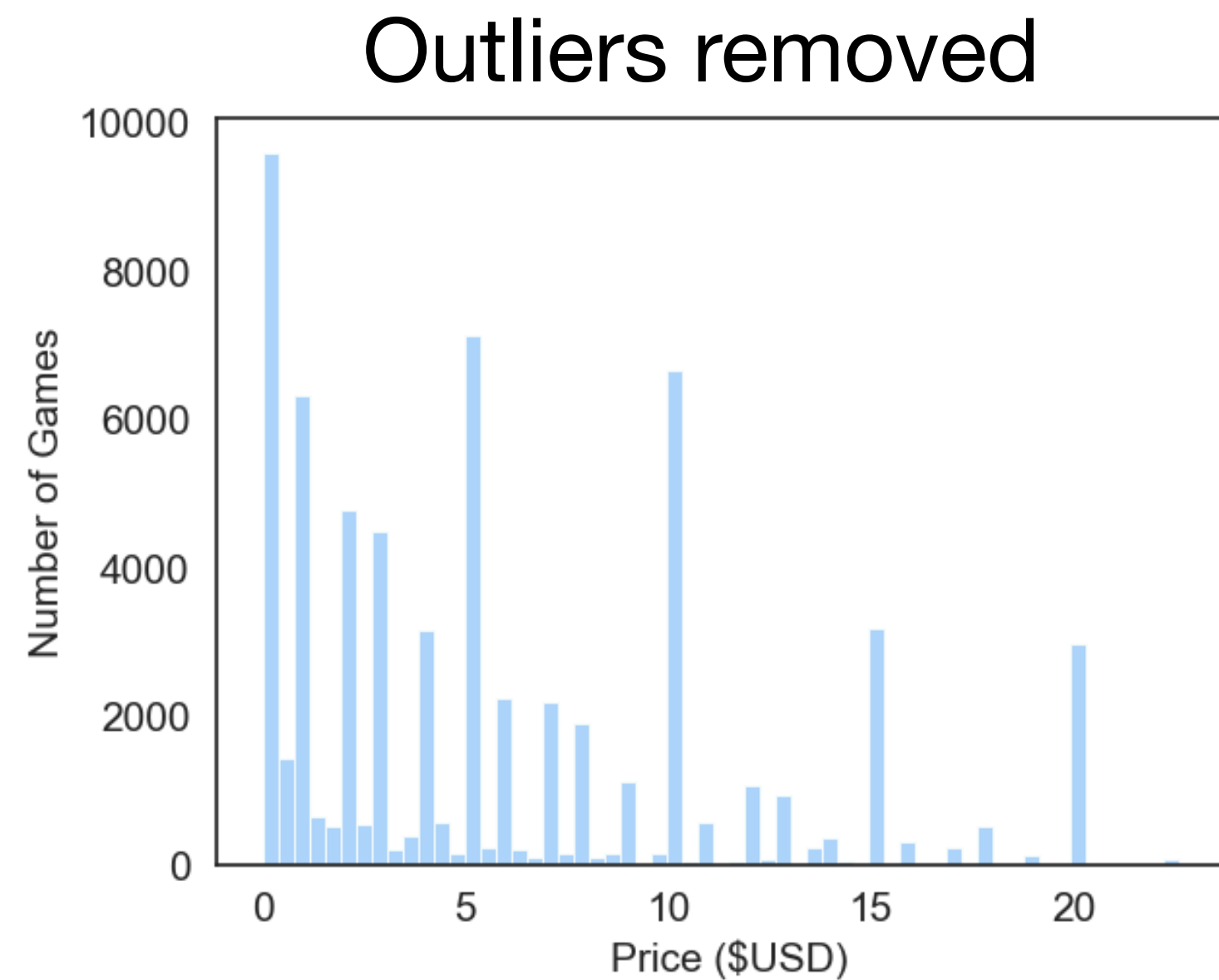
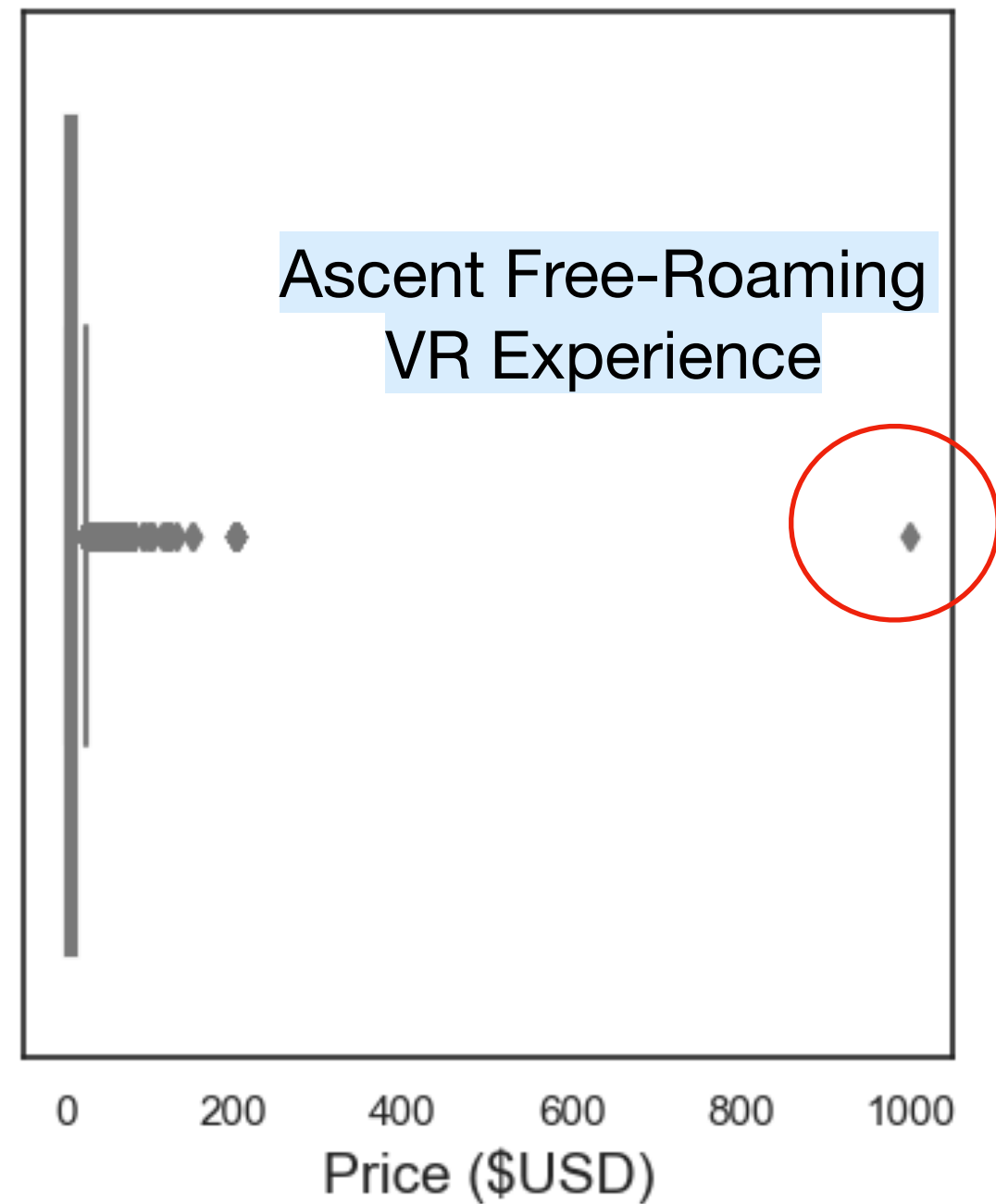
Steam is the largest digital distribution platform for PC games, with an estimated 75% of the market share in 2013 ([IHS Screen Digest](#))

## Preprocessing:

1. After data cleaning, **69528** games were included in the analysis
2. After feature engineering, the following variables were included:

*Name, Release date, Estimated # of Owners, Price, Developer, Publisher, Categories, Genres, Tags, Proportion Positive Reviews, Number of Total Reviews*

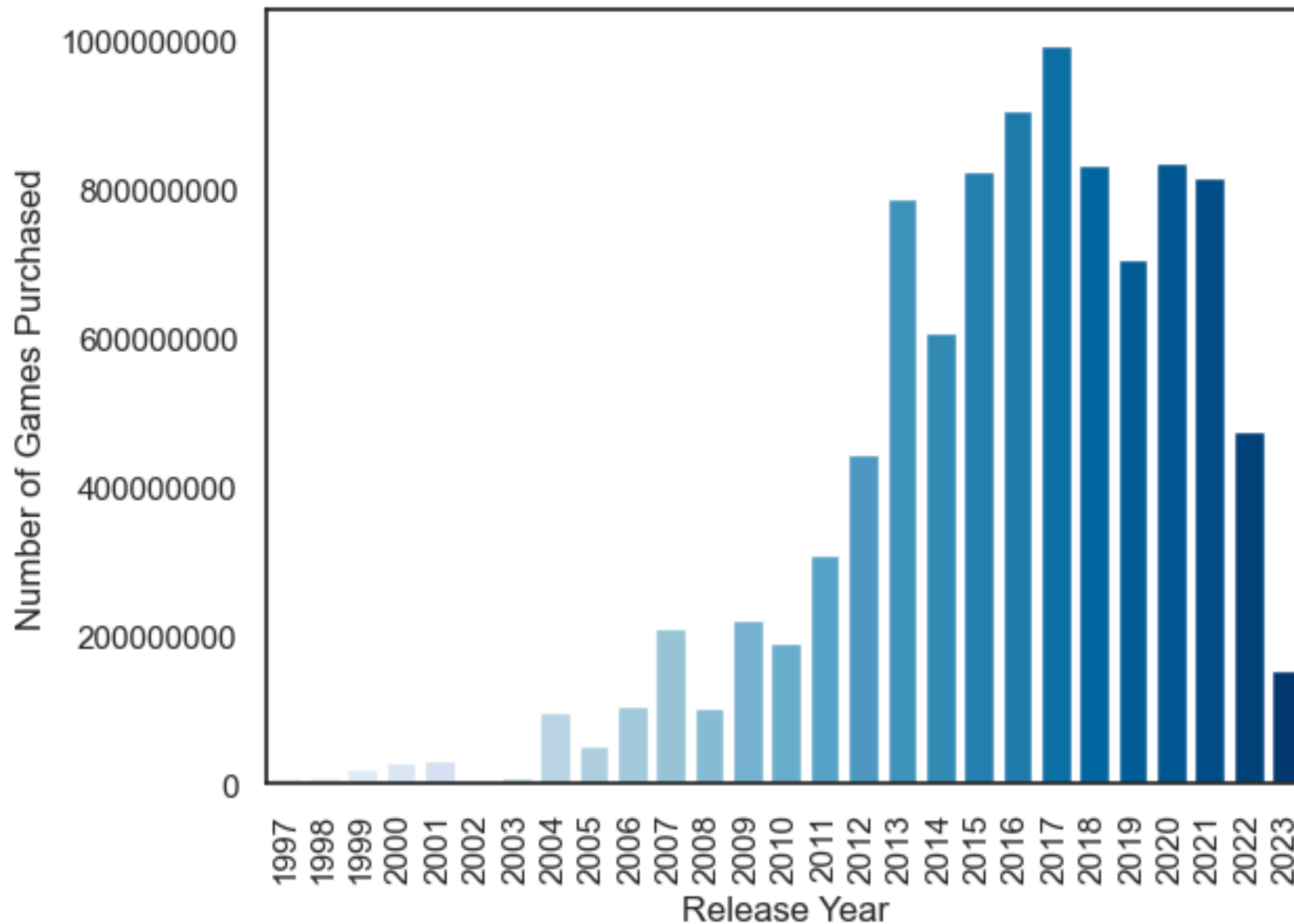
# Exploratory Analyses of Price & Market Size



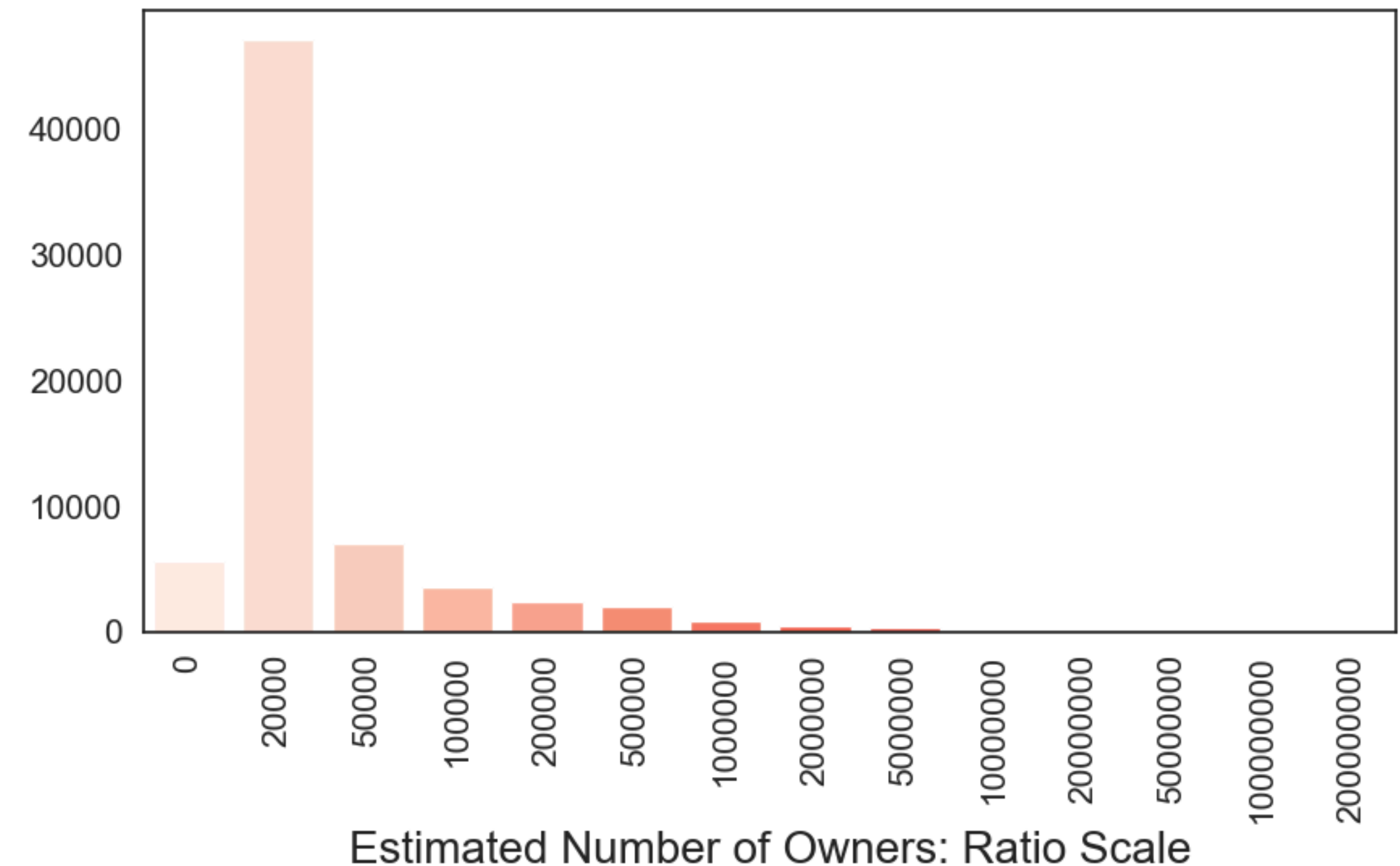
1. The most expensive game is a **VR simulation** that costs \$999
2. The most common price-point is free, and the average game costs **\$7.23 USD**
3. The number of Steam games released in **2022 is 11841**, and has increased year-by-year

# Exploratory Analyses of Price & Market Size

Estimated Number of Games Purchased by Release Year



Estimated Number of Game Owners by Category

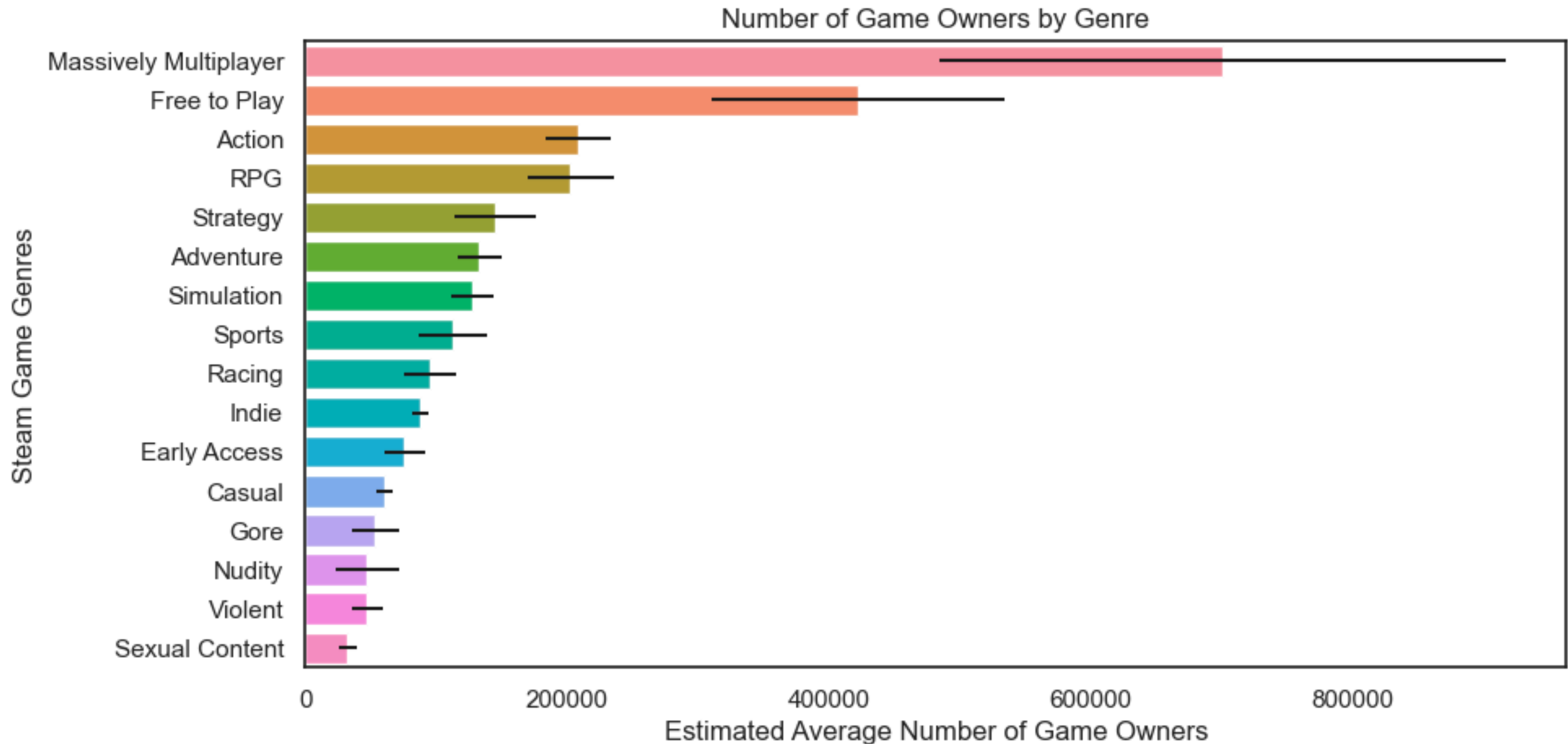


1. There is an estimated all-time **10 billion games sold** on Steam

2. **Most games sell 20,000 copies or fewer** on steam.  
Only a very small proportion sell more than 100,000 copies

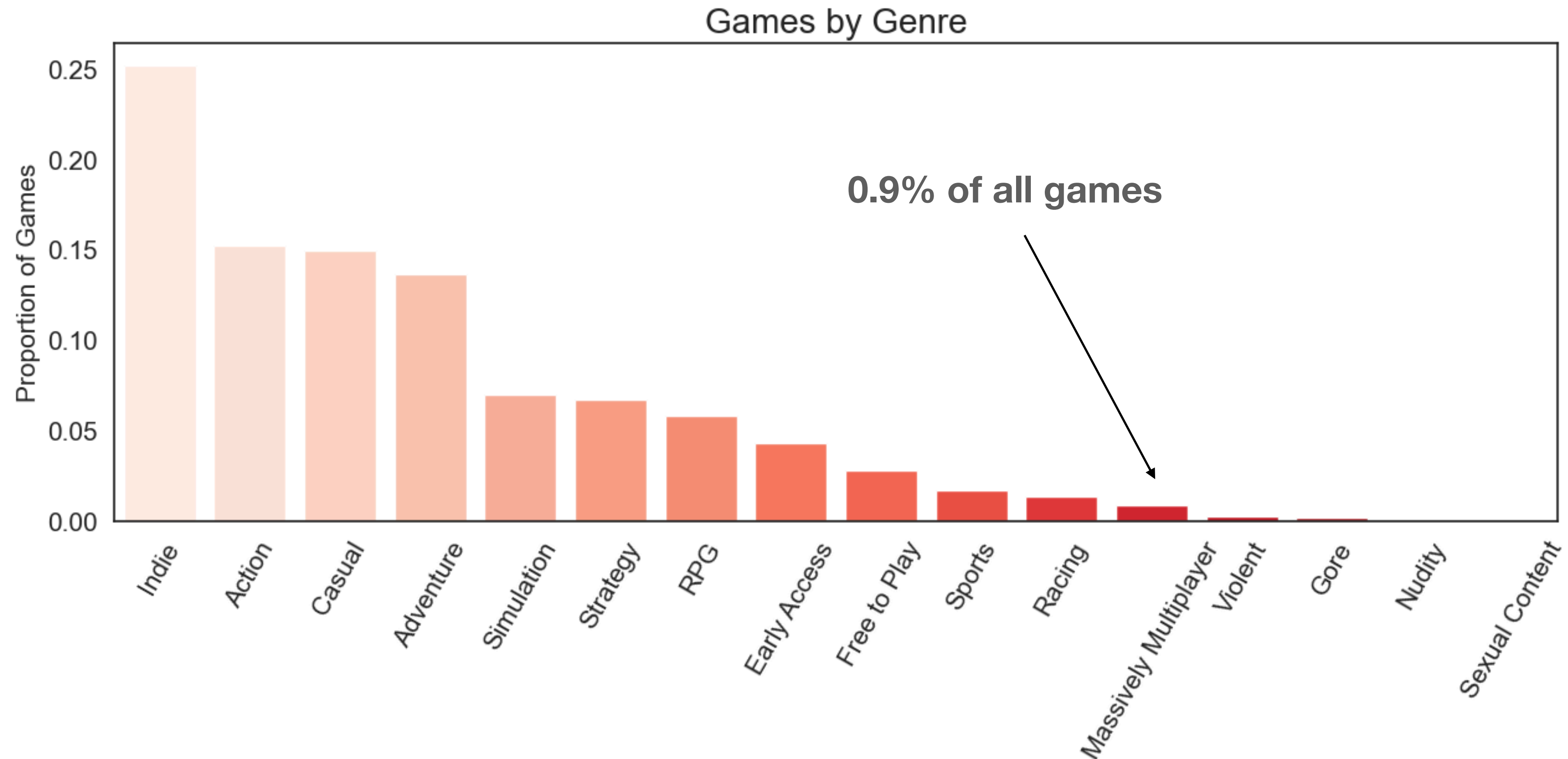


# The Largest Games are Massively Multiplayer



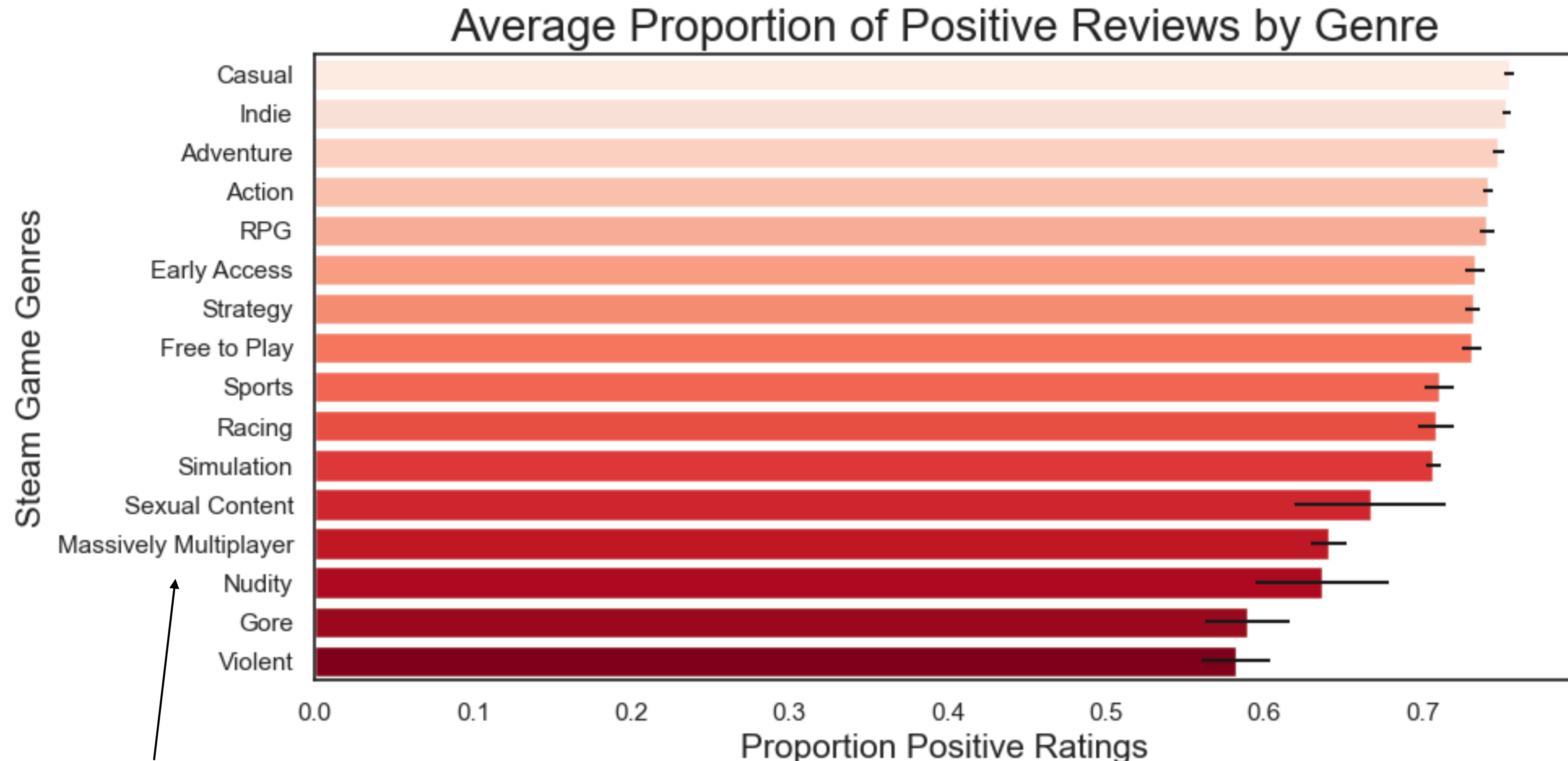
**Note:** A game can have multiple genres.

# Yet, Massive Multiplayer Games are Uncommon



1. The most common games are indie, with the least common games PG18+. This is likely because smaller developers build indie games, and do not have the resources for MMO-type games.

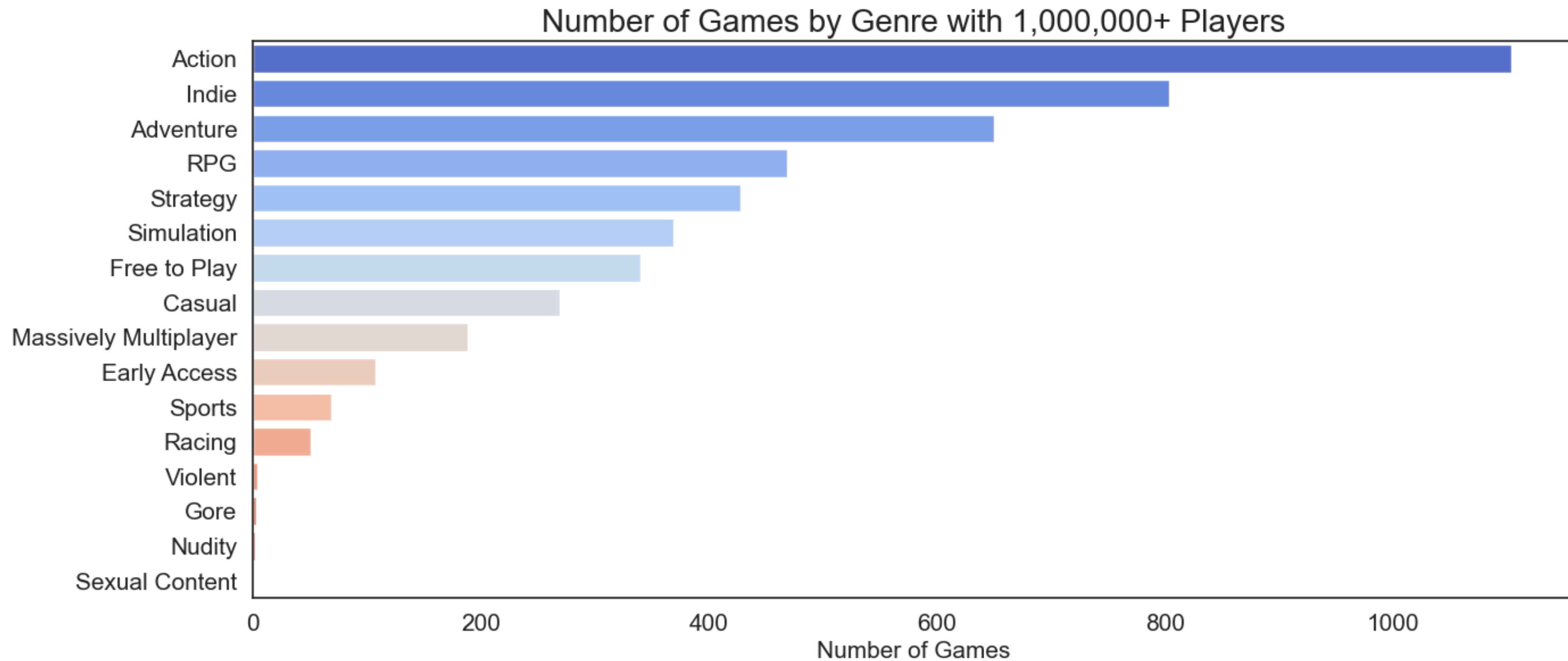
# Massive Multiplayer Games are Surprisingly Low Rated



Massive multiplayer games tend to be lower rated compared to other genres

Casual, Indie, Adventure, Action, and RPG game tend to have the highest ratings.

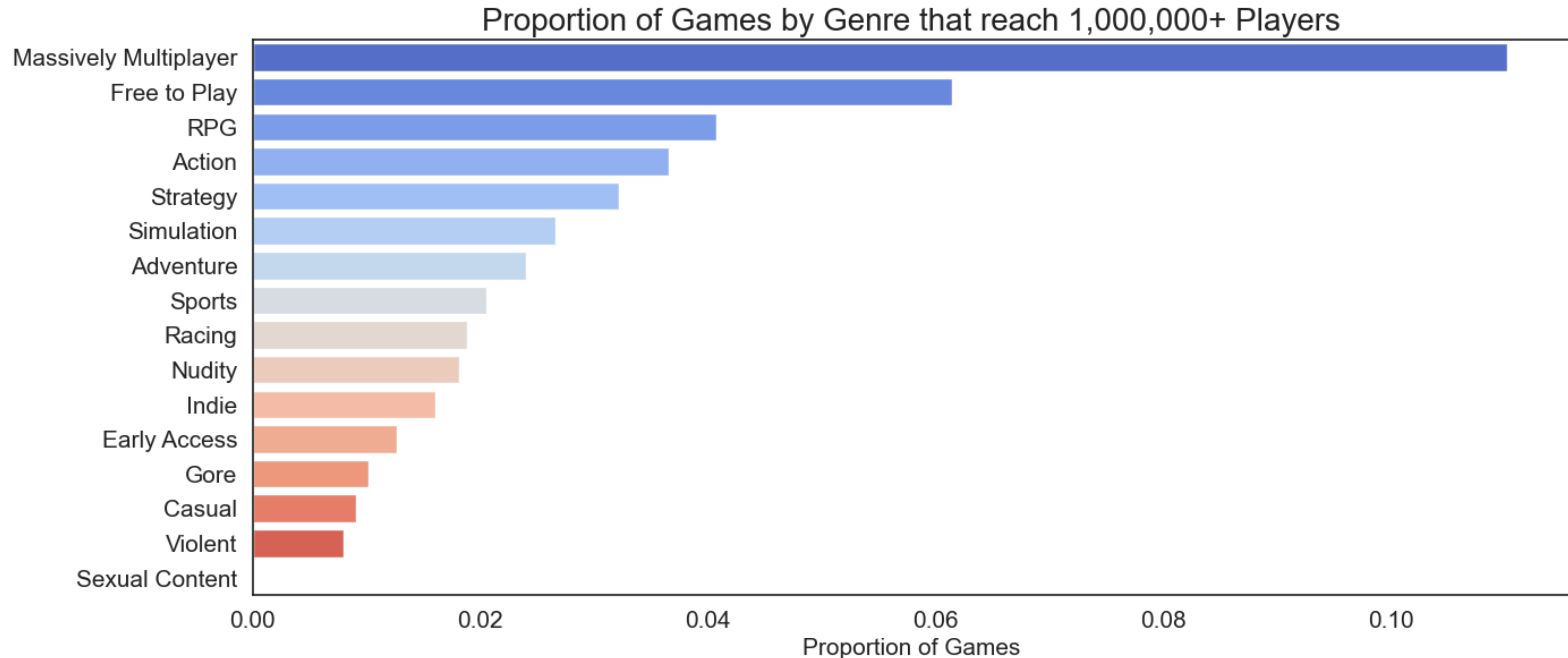
# Action, Indie, and Adventure are the Most Common Genres with 1 Million+ Purchases



**Over 1000 action games are highly successful**, whereas few sports, race, violent, or PG18+ games become highly successful.



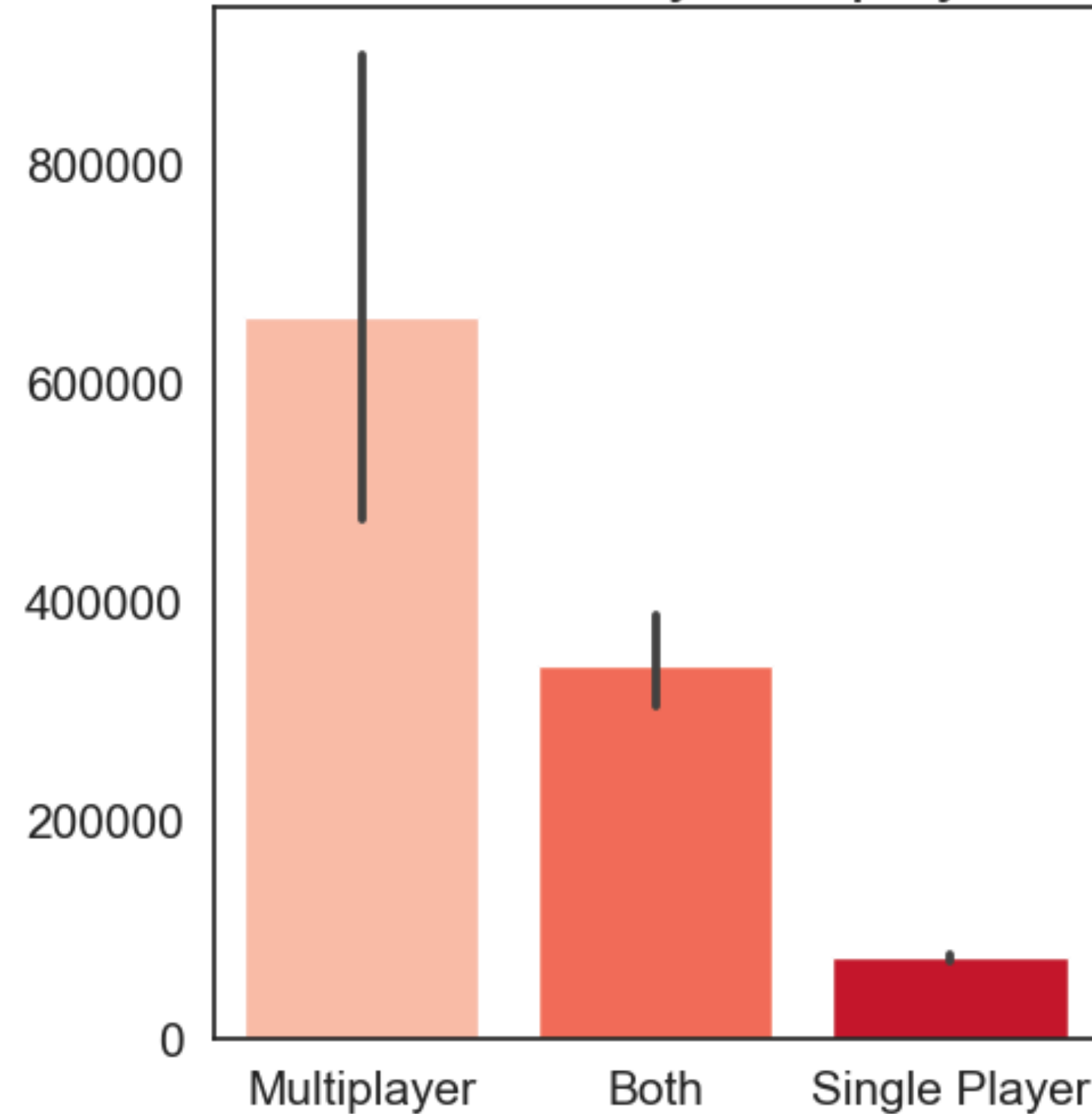
# Yet, Massively Multiplayer Games are Most Likely to be Highly Successful (>10%)



Over **1 in 10** massive multiplayer games become highly successful!  
The least successful genres are Sports, Racing, Violent, or PG18+ Games

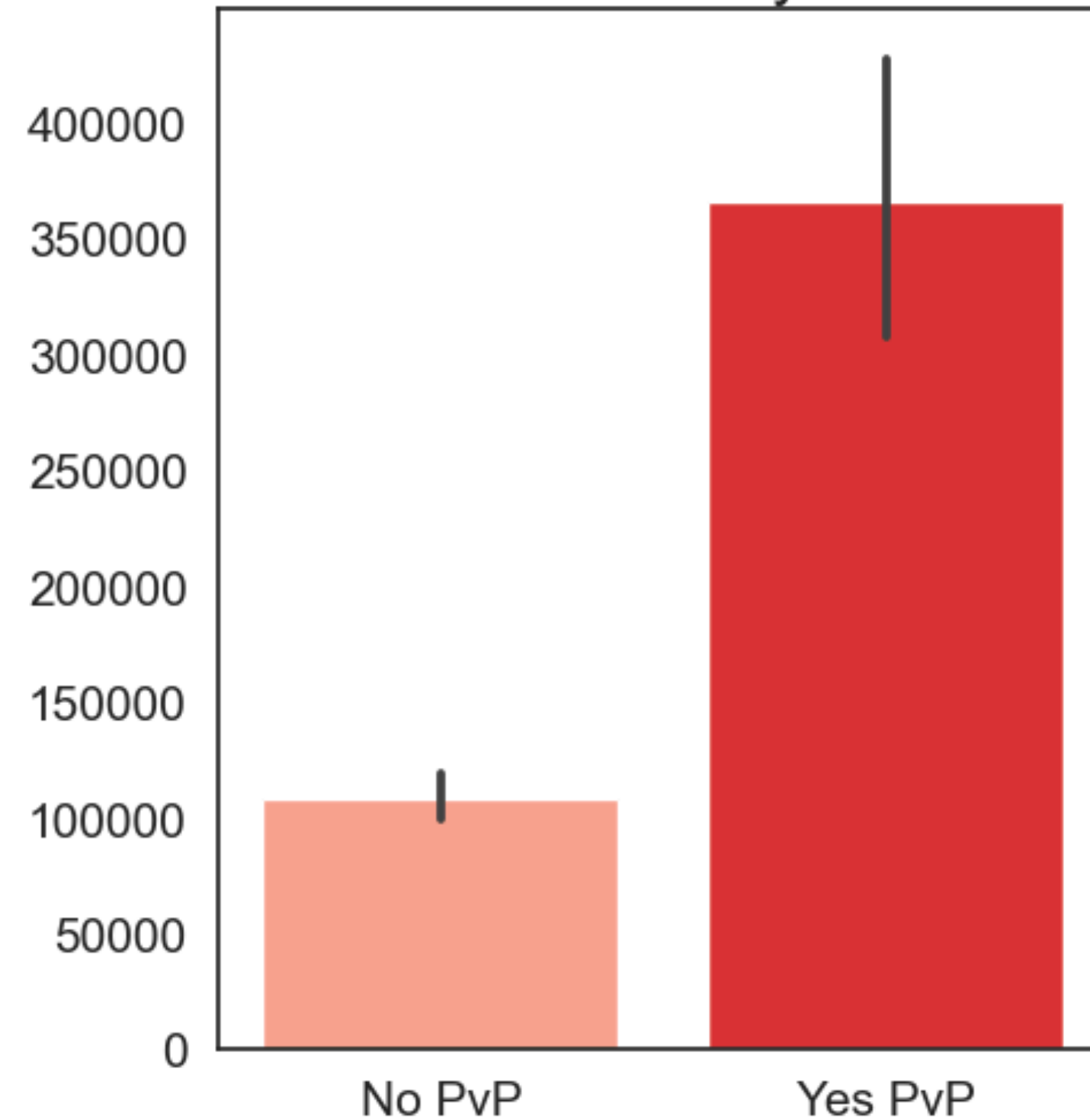
# Indeed, Multiplayer, PvP, and Co-op Games are Most Successful

Estimated Owners by Multiplayer Game



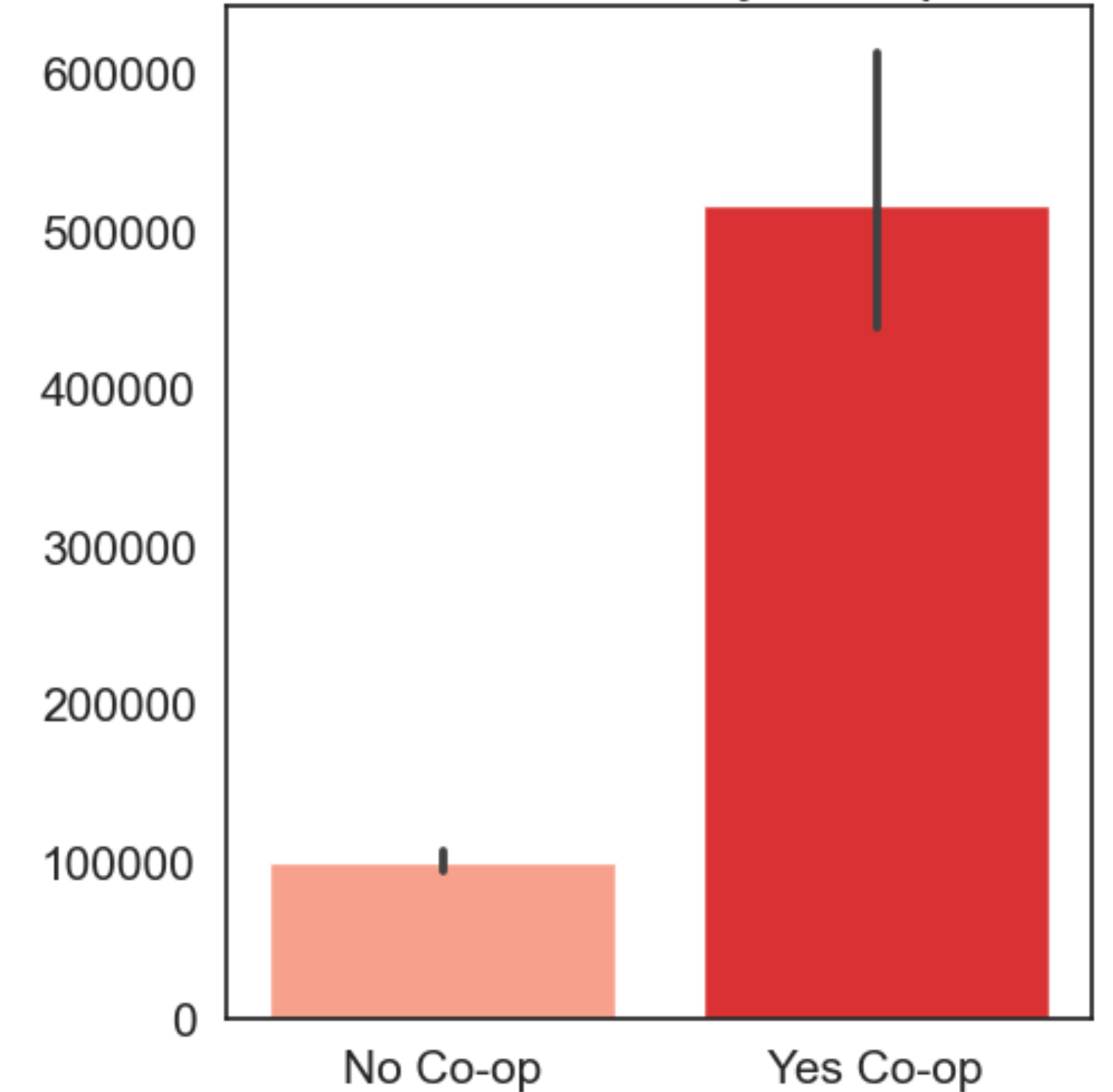
1. Multiplayer games tend to be much more successful compared to single-player games ( $p < 0.001$ )

Estimated Owners by PvP Game



2. Games with PvP are more successful than games without PvP ( $p < 0.001$ )

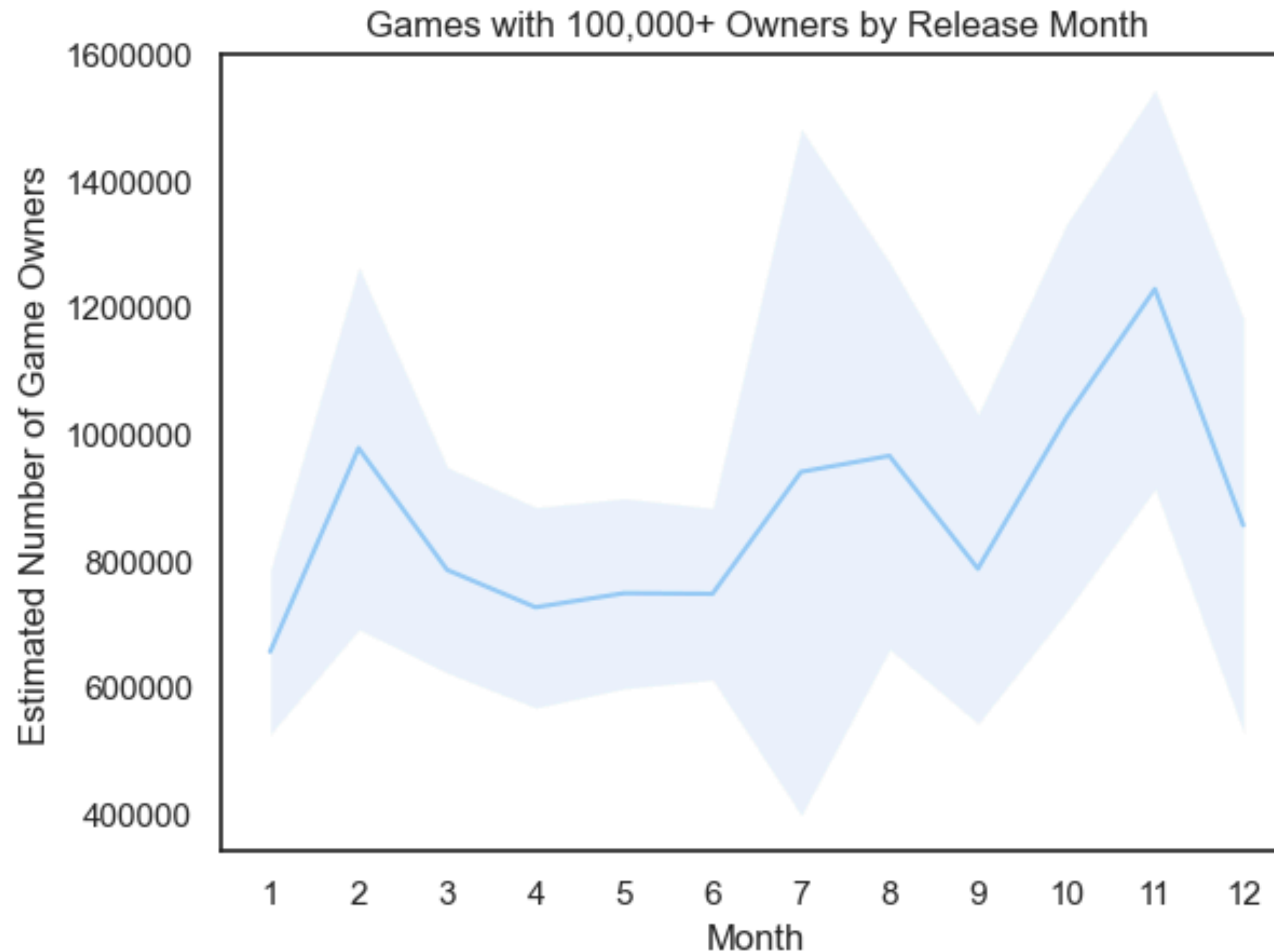
Estimated Owners by Co-op Game



3. Games with co-op options are more successful than games without co-op ( $p < 0.001$ )

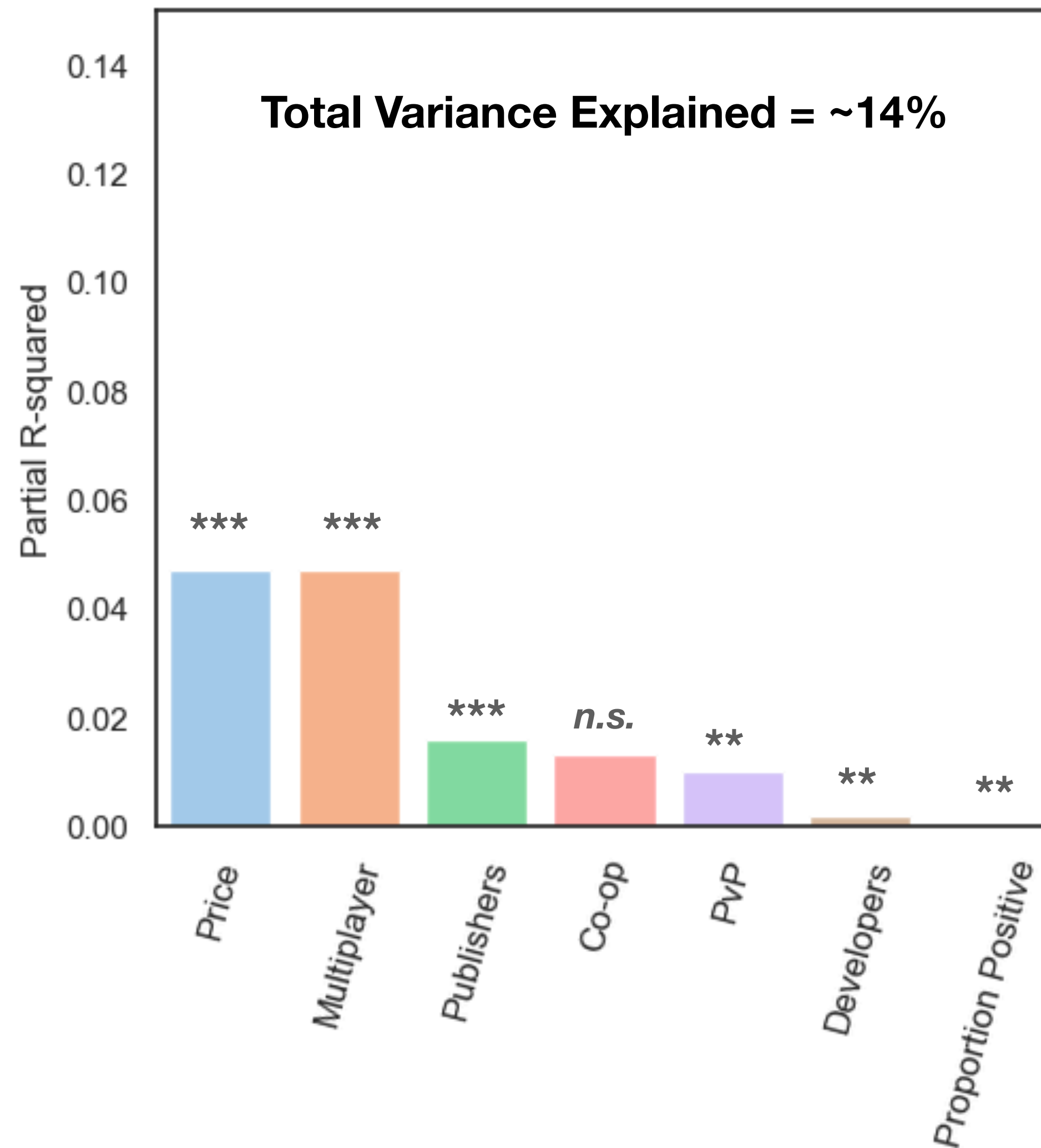
**Note:** Error bars reflect 95% CI around the mean

# On Average, November is the Best Time to Release a Game



**Note:** Shaded regions reflect 95% CI around the mean;  $n = 9710$

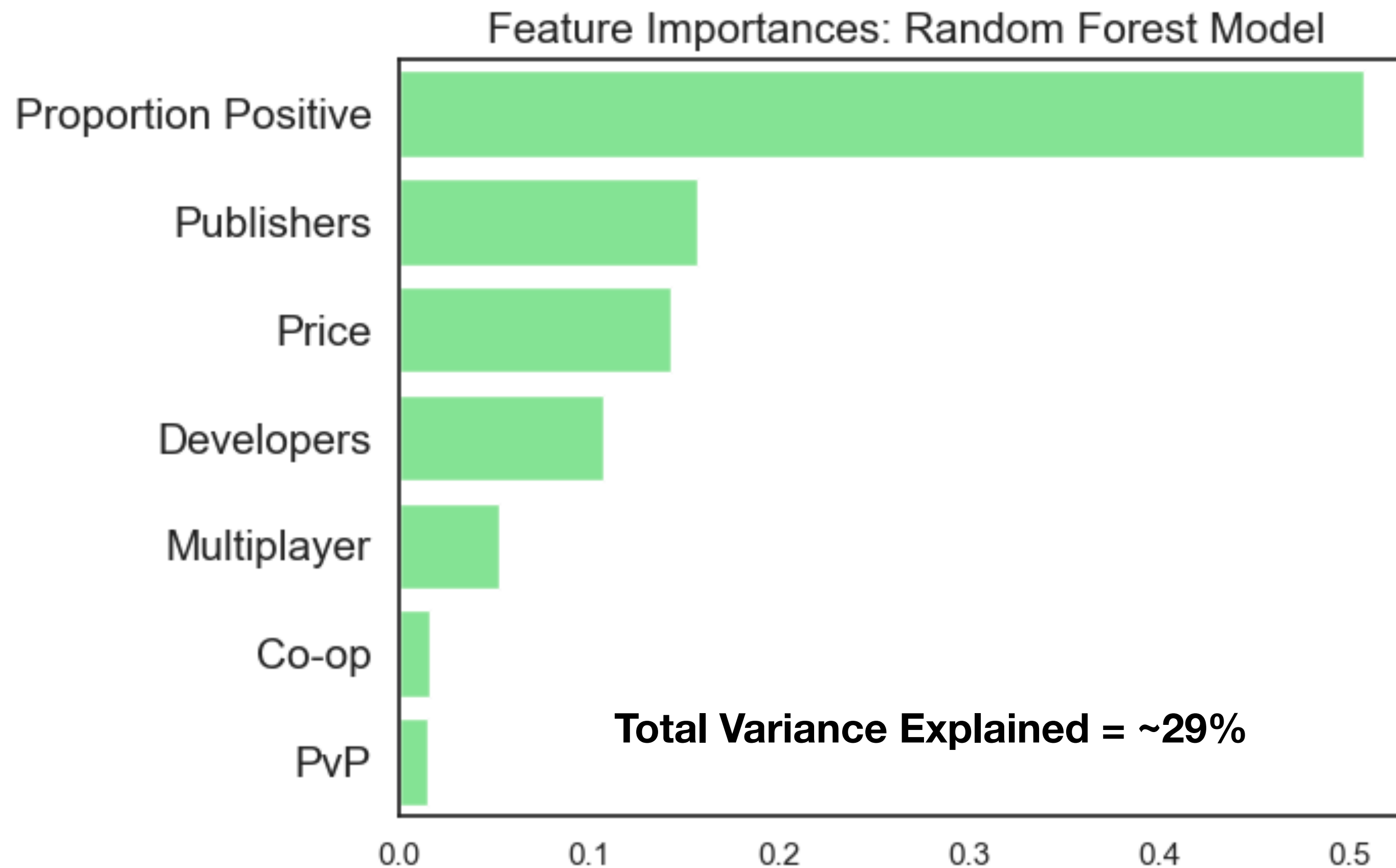
# Factors Predictive of Game Purchases: Regression



1. Surprisingly, a greater price point is moderately associated with number of game purchases (5% variance explained). This is likely due to **games with larger budgets** being priced higher.
2. **Multiplayer games** are moderately predictive of number of game purchases.
3. The number of games previously handled by a Publisher (“Publishers”) is weakly predictive of number of game purchases.
4. Only 14% of the variance is explained, suggesting that there is a better model.

**Note:** This model includes only moderately successful games (100,000+ estimated purchases;  $n = 13942$ )

# Factors Predictive of Game Purchases: Random Forest Model

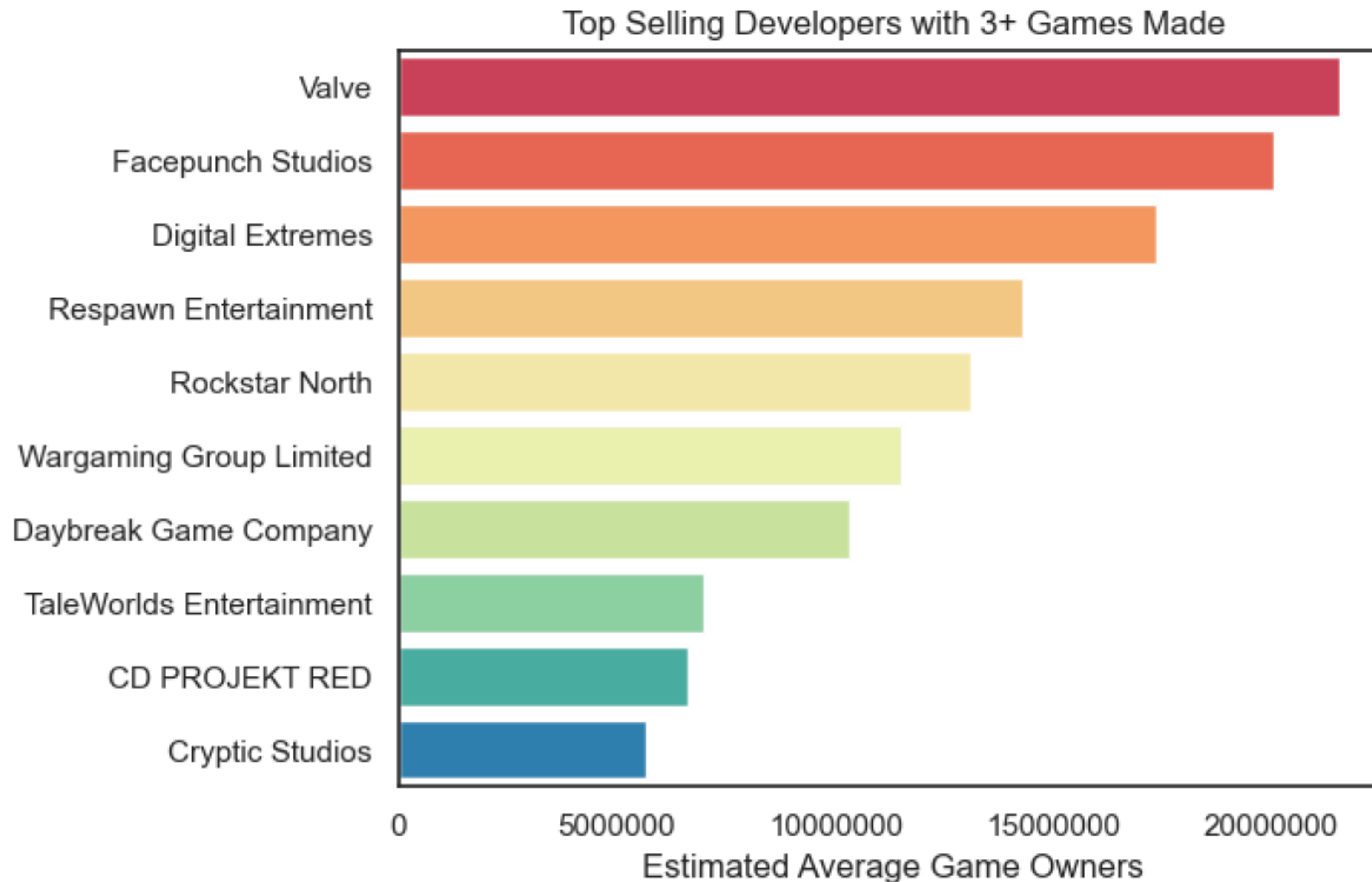


1. The random forest model indicates that “Proportion Positive” ratings is strongly related to game purchases.
2. Given that regressions can identify only \*linear relationships\*, this may indicate a complex relationship between positive ratings and number of purchase.
3. Number of games previously made by publishers and developers are also important factors here.
4. Variance explained is still fairly low.

**Note:** 1. This model includes only moderately successful games (100,000+ estimated purchases;  $n = 13942$ )  
2. Random forest models underestimate categorical predictors (e.g., multiplayer factor)

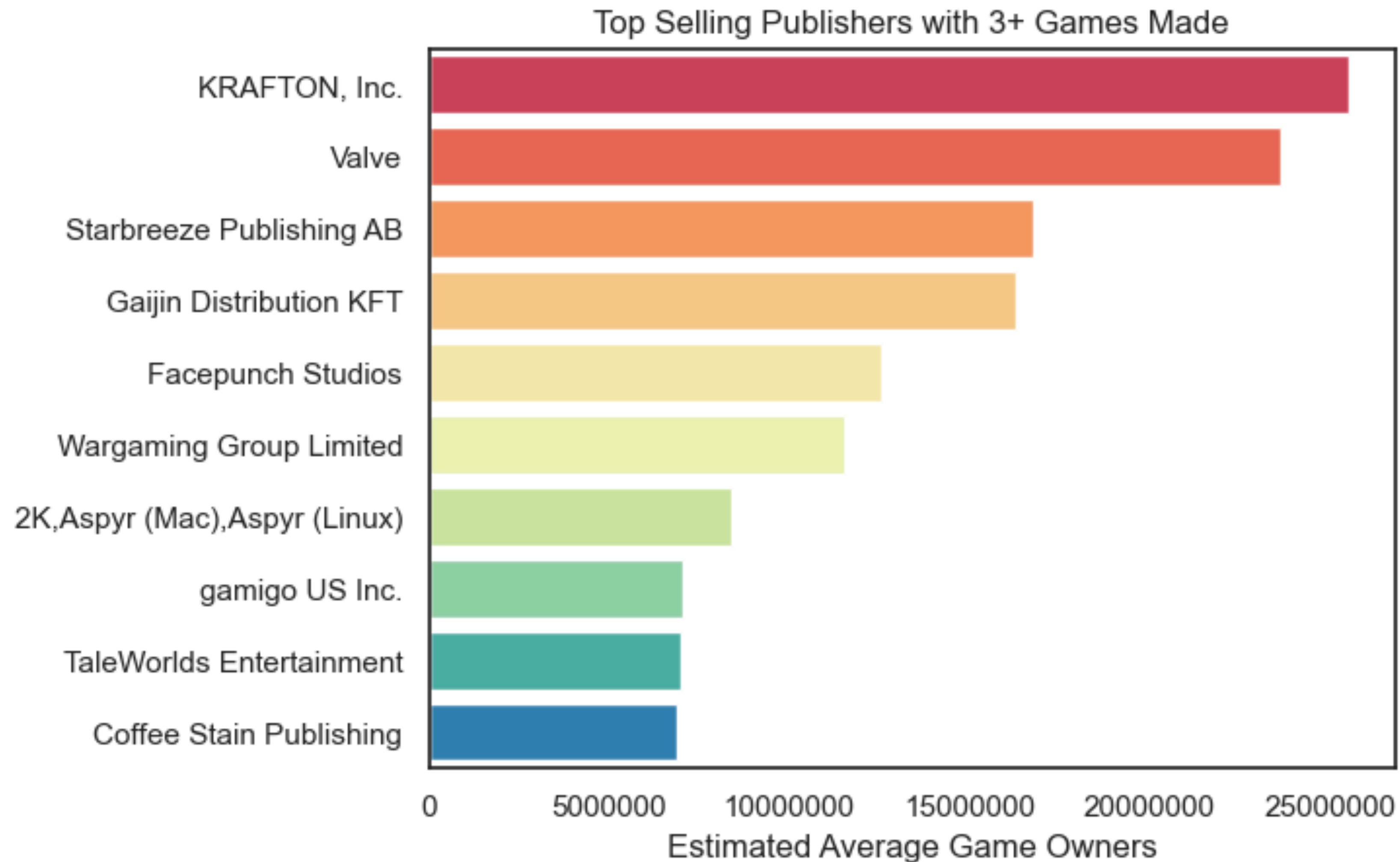


# Top Developers: Valve, Facepunch Studios, and Digital Extremes



*Future analyses should investigate factors associated with these developers*

# Top Publishers: KRAFTON, Inc., Valve, and Starbreeze Publishing



*Future analyses should investigate factors associated with these publishers*

# Summary of Findings:

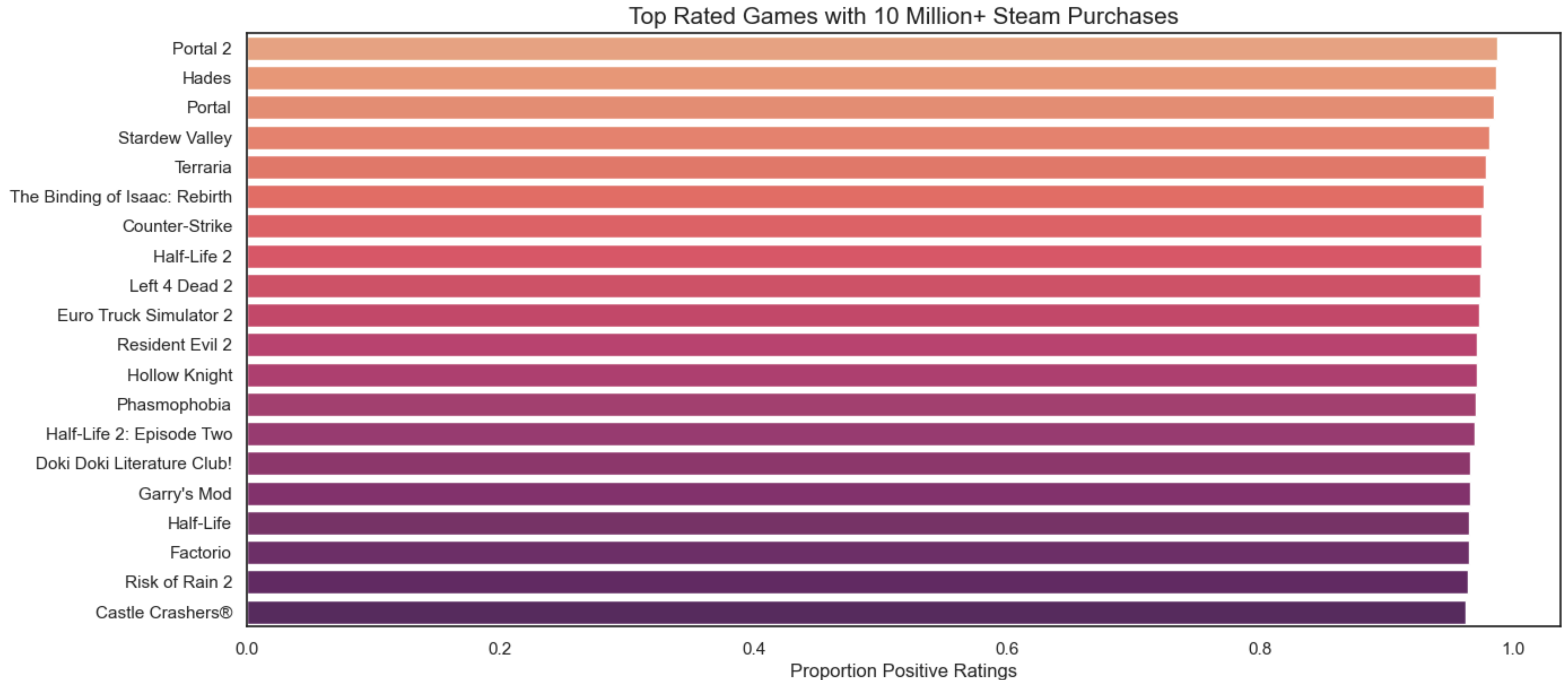
1. **Massive Multiplayer games are most popular**, but likely requires a high budget to make.
2. For smaller developers, **action and adventure games** are most likely to be successful.
3. Developers should add a **multiplayer or co-op** option to increase sales.
4. **November** is the best month to release a game.
5. A small proportion of games made by developers and publishers become highly successful.

# Future Investigations & Limitations:

1. These are a preliminary analyses of a Steam dataset due to limits in our data collected.
2. Future analyses should investigate factors associated with developer/publisher success.
3. Positive ratings have a complex relationship with sales. Need to better understand this factor.
4. With additional user data, we can build more targeted recommendations by genre.
5. Our models account for only ~30% of the variance. Need to understand the factors associated with the game itself that contributes to success: **What makes a game good?**

# User Recommendations: Filter by Type

We can filter our dataset by ratings, by genre, or by most successful developers. **Example:**





# User Recommendations: Game Similarity

Alternatively, we can recommend games based on their similarity. Here, I calculated the cosine similarity between user-rated game tags, after filtering by top games — see the Github for code.

## Example: “Slay the Spire”

```
game_recommender("Slay the Spire", 0.85, 50, 10)
```

|                         | Price | Proportion_Positive | Total_Reviews | Similarity |
|-------------------------|-------|---------------------|---------------|------------|
| Blood Card 2: Dark Mist | 5.99  | 0.863208            | 212           | 0.900000   |
| Dicey Dungeons          | 14.99 | 0.902158            | 8340          | 0.850000   |
| Iris and the Giant      | 17.99 | 0.888778            | 998           | 0.800000   |
| Roguebook               | 24.99 | 0.850710            | 2465          | 0.800000   |
| Across the Obelisk      | 19.99 | 0.941207            | 1956          | 0.750000   |
| Card Quest              | 9.99  | 0.881295            | 278           | 0.750000   |
| Poker Quest             | 14.99 | 0.926554            | 354           | 0.750000   |
| Legend Creatures(传奇生物)  | 4.99  | 0.909014            | 3539          | 0.726722   |
| Erannorth Reborn        | 19.99 | 0.863372            | 344           | 0.700000   |
| Gordian Quest           | 19.99 | 0.918421            | 4180          | 0.700000   |
| Knock on the Coffin Lid | 24.99 | 0.863158            | 475           | 0.700000   |

## Example: “The Forest”

```
game_recommender("The Forest", 0.85, 50, 10)
```

|                  | Price \$USD | Proportion_Positive | Total_Reviews | Similarity |
|------------------|-------------|---------------------|---------------|------------|
| The Long Dark    | 34.99       | 0.914913            | 90284         | 0.80       |
| Green Hell       | 24.99       | 0.870716            | 47856         | 0.75       |
| Subnautica       | 29.99       | 0.963634            | 199033        | 0.75       |
| Project Zomboid  | 19.99       | 0.923807            | 109328        | 0.70       |
| The Infected     | 13.99       | 0.879056            | 4407          | 0.70       |
| Volcanoids       | 19.99       | 0.858340            | 4772          | 0.70       |
| Medieval Dynasty | 34.99       | 0.906533            | 26373         | 0.65       |
| Raft             | 19.99       | 0.932132            | 196130        | 0.65       |
| Satisfactory     | 29.99       | 0.971783            | 100683        | 0.65       |
| Darkwood         | 14.99       | 0.946186            | 13175         | 0.65       |
| Don't Be Afraid  | 9.99        | 0.859043            | 376           | 0.65       |