

Steam Game Popularity Analysis & Prediction

Allen Arriaga

11/13/2025

Video game popularity and player engagement are critical factors for success in the gaming industry. On platforms like Steam, pricing strategies such as discounts can significantly influence player activity and overall game sales. Understanding the factors that drive game popularity helps developers and publishers make more informed marketing and pricing decisions. This analysis explores how discounts, developer presence, and player behavior interact to shape a game's overall visibility and performance on the platform.

In this analysis, I used data analysis and machine learning techniques to examine how discounts, pricing, and player activity relate to a game's popularity and engagement levels on Steam. By applying Random Forest regression modeling, I investigated which features—such as discount percentage, concurrent users, and reviews—most strongly predict game ownership and positive player feedback.

The main question of this investigation is: "How do discounts and player activity affect game popularity on Steam?"

1. Data Description and Processing

The dataset used in this analysis was collected from the Steam gaming platform and includes a variety of game-specific, developer, and player engagement characteristics. Each observation represents data from an individual game available on Steam during the study period.

This dataset contains several types of variables related to game pricing, player activity, and popularity metrics. Key features include game price, discount rate, number of owners, concurrent users, developer, and counts of positive and negative reviews. These variables provide insight into how economic and engagement factors interact to influence overall game success.

Variable Descriptions

appid: The unique Steam Application ID for the game.

name: The name of the game.

developer: A comma-separated list of the game's developers.

publisher: A comma-separated list of the game's publishers.

owners: The estimated number of owners of this application on Steam.

average_forever: The average total playtime since March 2009 (minutes).

median_forever: The median total playtime since March 2009 (minutes).

average_2weeks: The average playtime in the last two weeks (minutes).

median_2weeks: The median playtime in the last two weeks (minutes).

ccu: The peak Concurrent Users (CCU) for the game yesterday.

price: The current US price of the game.

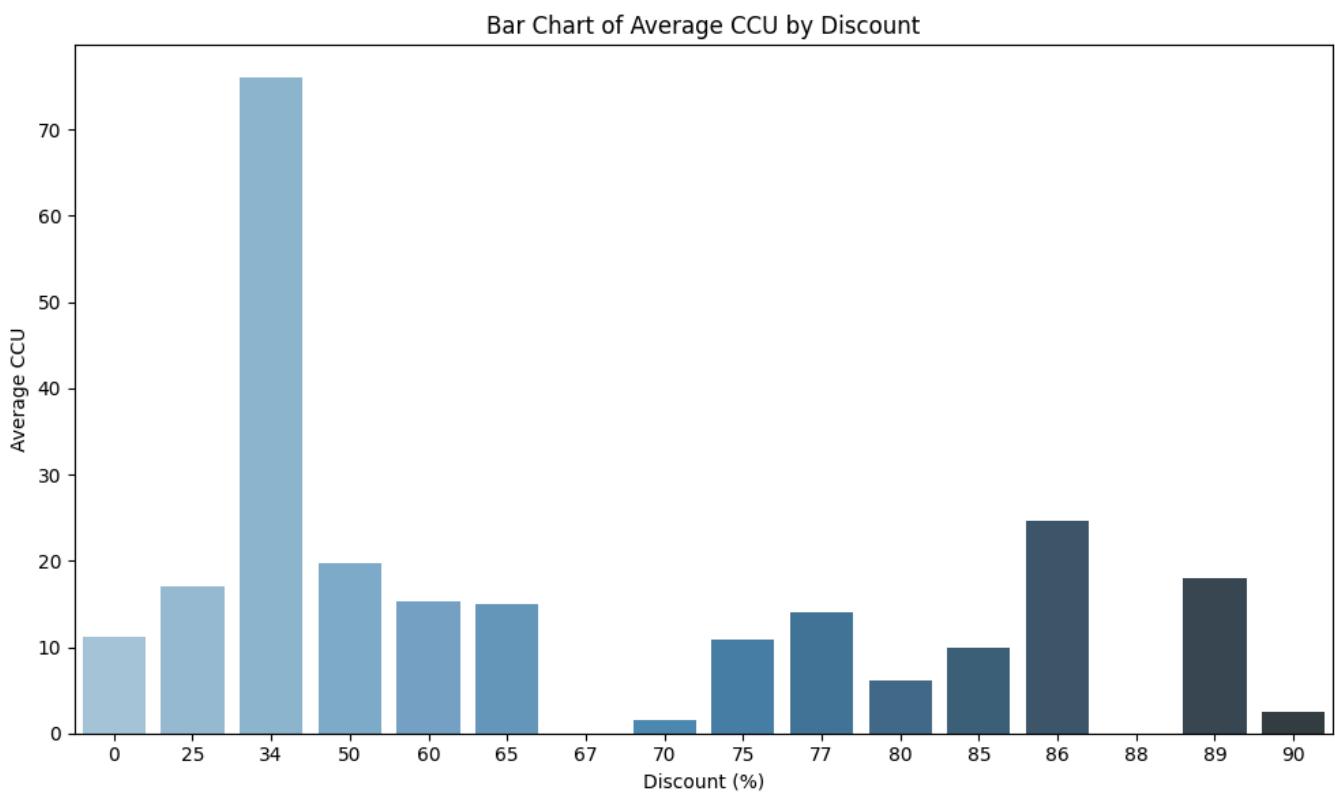
initialprice: The original US price of the game.

discount: The current discount on the game (percent).

The dataset was preprocessed in MySQL before being imported into Python for analysis. Initially, the dataset contained 347 records. After preprocessing, the cleaned dataset contained 339 complete observations. Additional outlier filtering was then performed in Python using the interquartile range (IQR) method, resulting in a final dataset of 270 games used for modeling and visualization.

2. Steam Game Trends and Insights

Do higher discounts bring in more players?



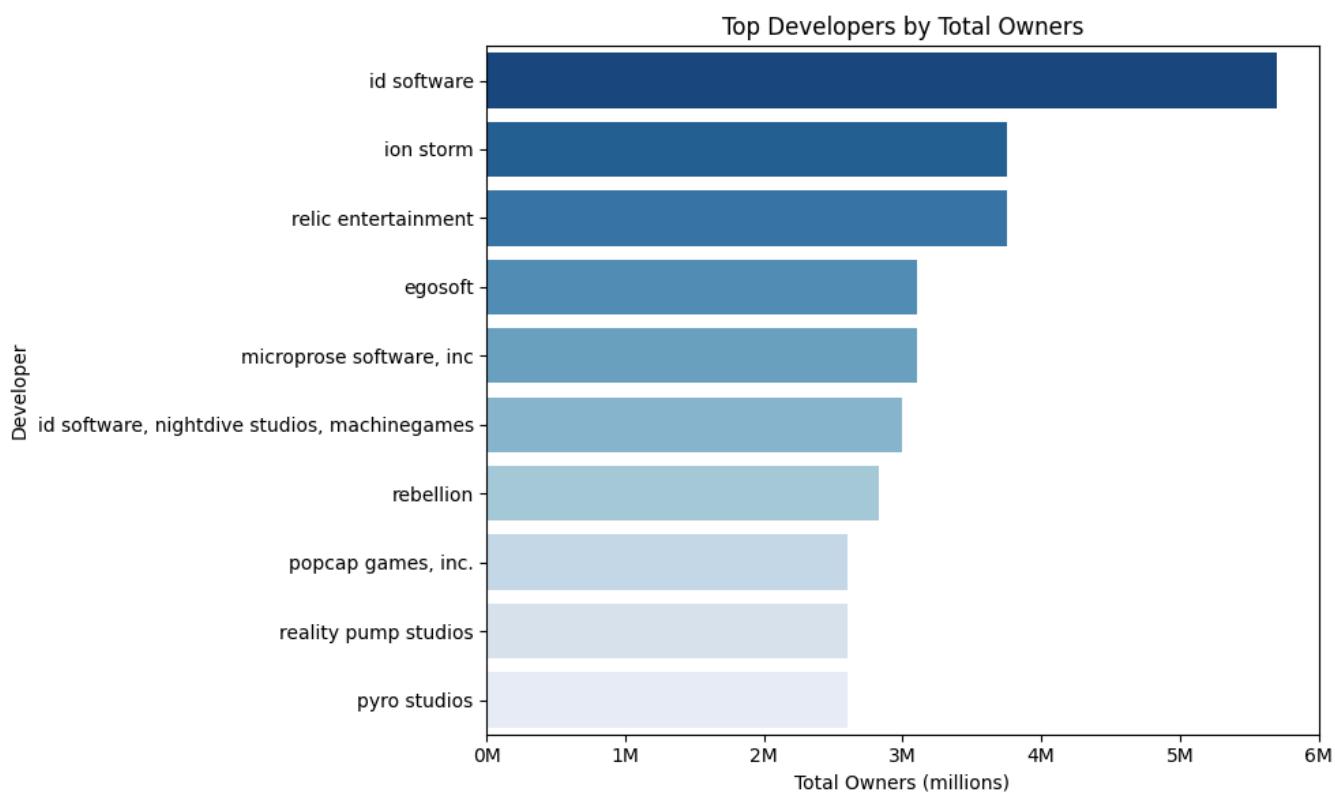
Moderate discounts seem to draw the most active players, but extreme discounts don't necessarily lead to higher player activity — suggesting that game popularity matters more than price cuts alone.

What developers make popular games (based on the number of owners)?

	Developer	Total Owners
	id software	5,700,000
	ion storm	3,750,000
	relic entertainment	3,750,000
	egosoft	3,100,000
	microprose software, inc	3,100,000
id software, nightdive studios, machinegames		3,000,000
	rebellion	2,825,000
	popcap games, inc.	2,600,000
	reality pump studios	2,600,000
	pyro studios	2,600,000

number of times a specific developer shows up in the dataset

	Developer	Game Count
	popcap games, inc.	18
	id software	9
	egosoft	6
	rebellion	5
	microprose software, inc	5
	pyro studios	4
	reality pump studios	3
	relic entertainment	3
	ion storm	3
id software, nightdive studios, machinegames		2

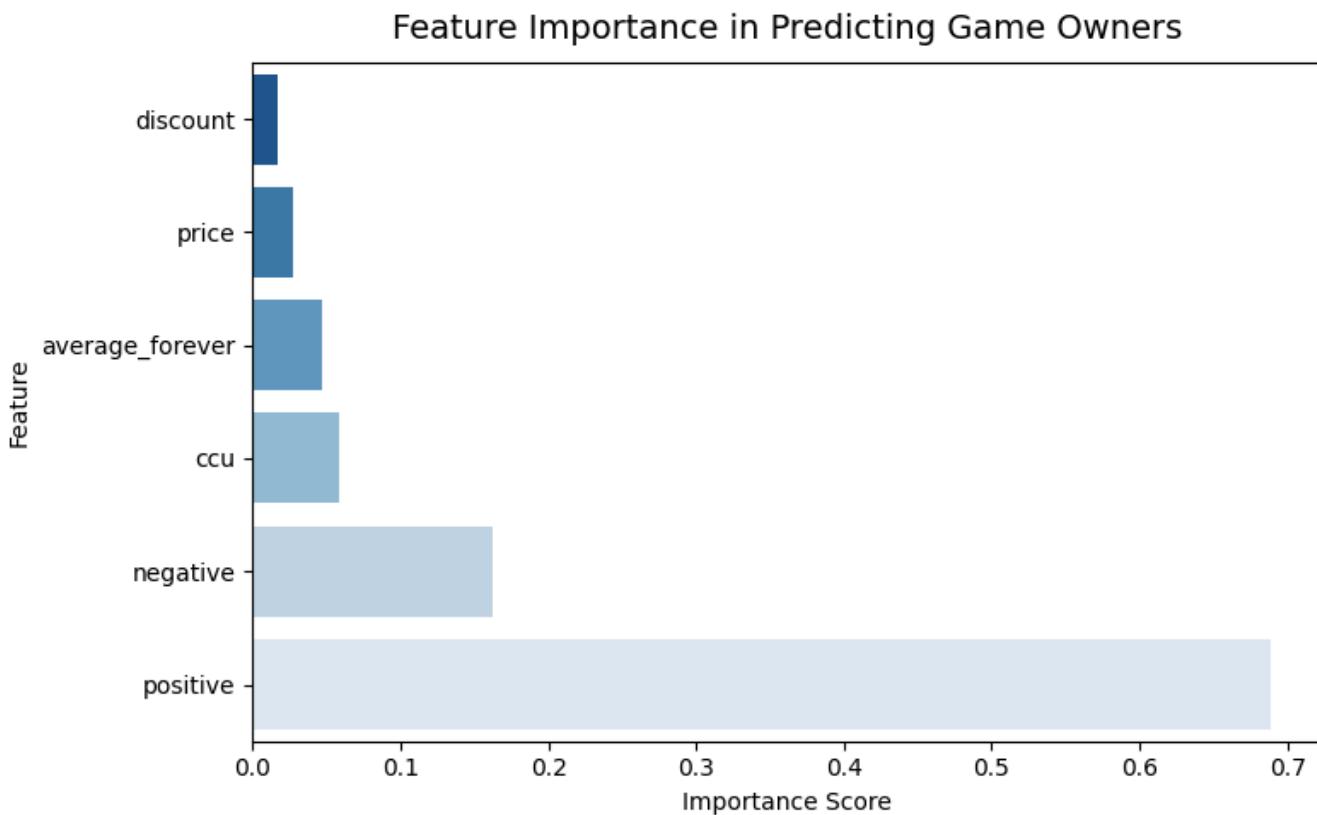


3. Predictive Modeling for Game Success

how many good reviews will a new game have?

To answer this question, I trained a Random Forest model using features like price, discount, positive and negative reviews, playtime, and concurrent users to predict the number of owners. The model learns patterns from historical game data and evaluates its accuracy using the R² score. Random Forest model was best for this analysis due to its ability to capture the complex relationships between game features—like price, reviews, and player activity—and the number of owners, giving accurate and reliable predictions.

```
R2 score: 0.5483477182099525
positive      0.689091
negative      0.162119
ccu          0.058126
average_forever 0.047186
price         0.027191
discount       0.016287
dtype: float64
```



The feature importance results indicate that positive reviews(positive), negative reviews, and ccu are the strongest predictors of game ownership. In contrast, price and discount rate have minimal influence. This suggests that player engagement and community feedback are far more indicative of a game's success than short-term pricing factors.

Predicting how many owners a new game would get given hypothetical inputs

Predicted number of owners: 1275000

R² score: 0.8614432198138235

```
ccu          0.318455
owners      0.316578
negative    0.259790
average_forever 0.050455
price        0.044653
discount     0.010069
dtype: float64
```

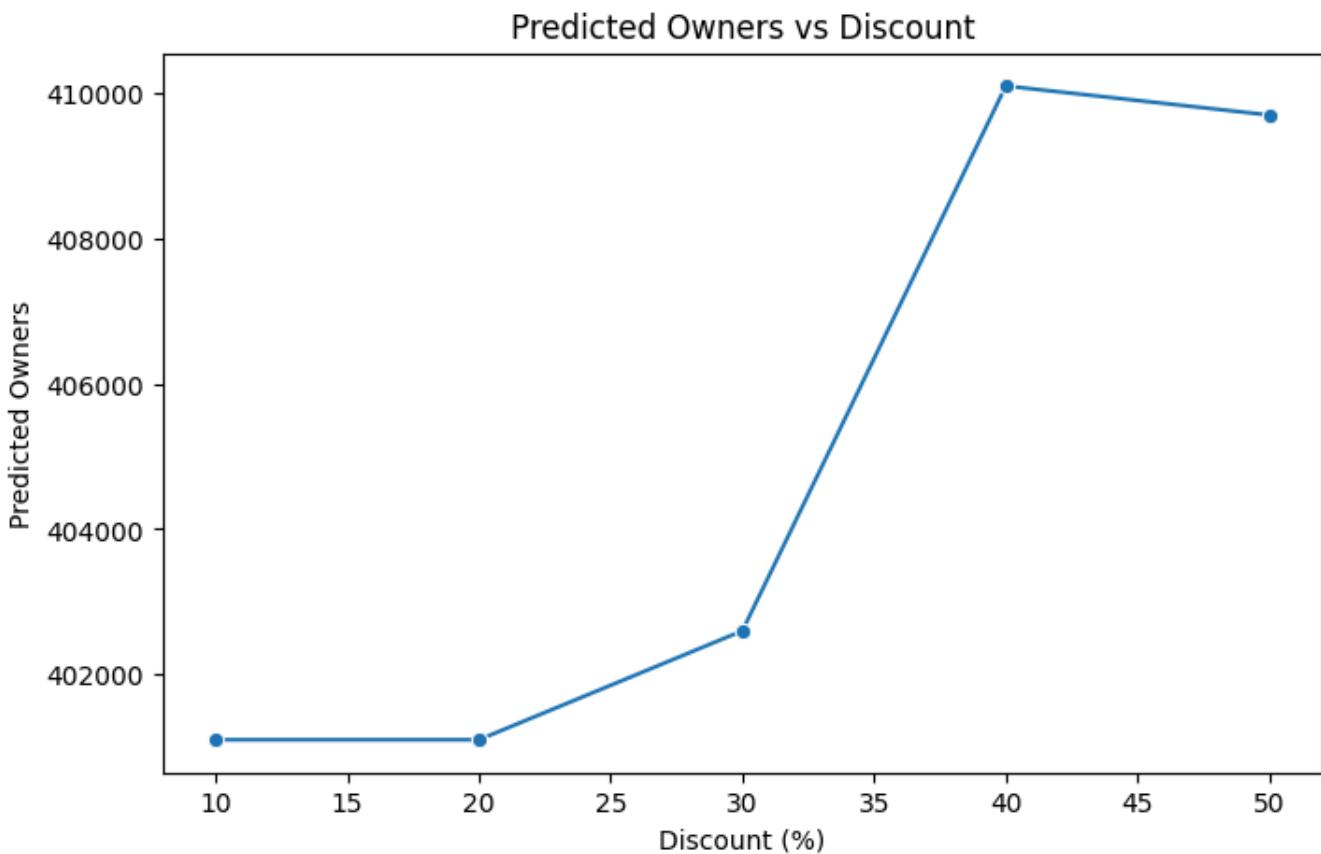
Predicted number of positive reviews: 14251

This Random Forest model is a good fit($R^2 = 0.86$), effectively capturing the relationship between gameplay and popularity factors. Positive review counts are primarily driven by engagement (CCU) and player reach (owners), suggesting that active communities and game visibility are far more important predictors of positive reception than pricing or discounts.

how does predicted ownership change as discount increases?

	price	discount	positive	negative	average_forever	ccu	\
0	9.99	10	500	800	2000	1500	
1	9.99	20	500	800	2000	1500	
2	9.99	30	500	800	2000	1500	
3	9.99	40	500	800	2000	1500	
4	9.99	50	500	800	2000	1500	

	predicted_owners
0	401100.0
1	401100.0
2	402600.0
3	410100.0
4	409700.0



The table and graph show that as the discount increases from 10% to 40%, the predicted number of owners rises from about 401,100 to 410,100. However, at 50%, the predicted ownership slightly drops to 409,700, suggesting that while moderate discounts drive higher ownership, very large discounts may not continue this trend. The increase is gradual at first and becomes more noticeable at higher discounts, but the small dip at 50% indicates that beyond a certain point, deeper discounts might not yield additional growth in ownership.

4. Conclusion

This analysis explored how discounts, pricing, and player activity influence game popularity and engagement on Steam. The results show that while discounts can modestly increase game ownership, their effect is limited compared to factors like player engagement and community feedback. Random Forest modeling revealed that positive and negative reviews, as well as concurrent player counts (CCU), are the strongest predictors of game popularity, while price and discount rate have minimal impact.

Overall, the findings suggest that successful games are driven more by sustained player interest and reputation than by temporary pricing strategies. Developers aiming to boost popularity should prioritize improving gameplay experience and fostering active player communities rather than relying heavily on steep discounts.