Predicting Continued Popularity of Games on Twitch

Daniel Ansted

Executive Summary

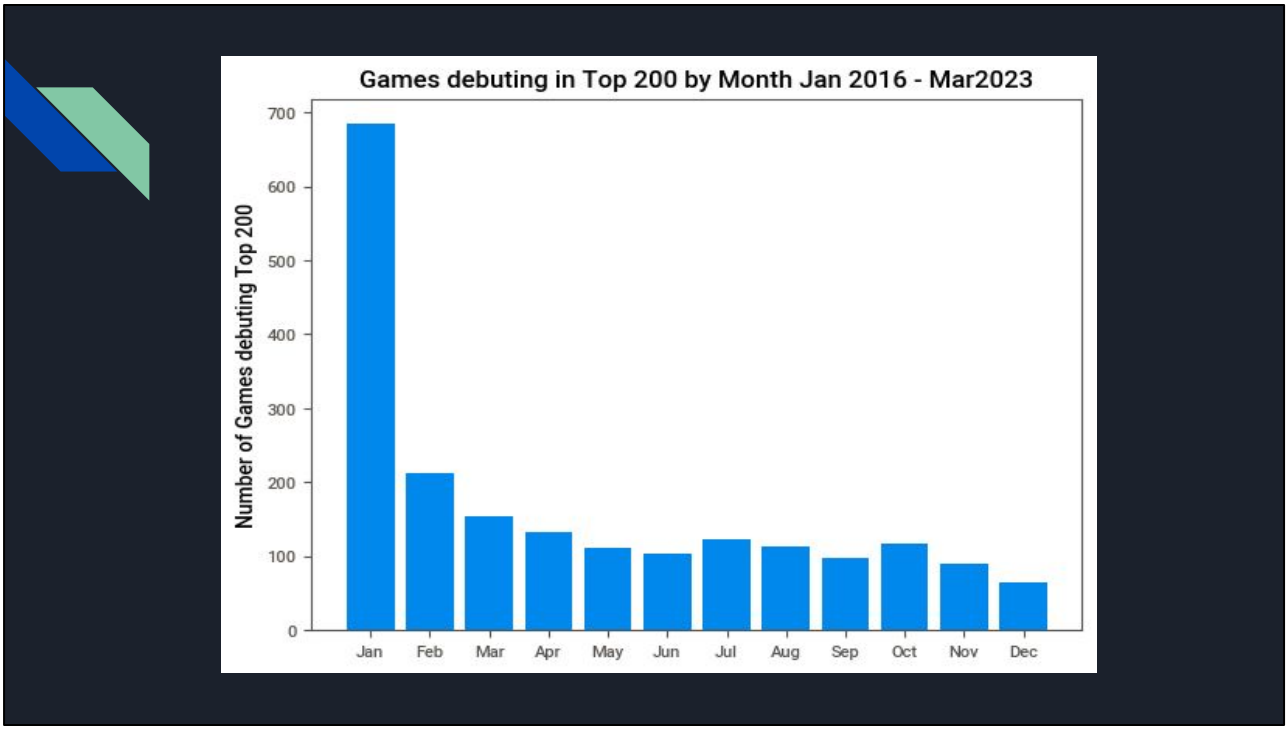
- Our machine learning model could allow us to target an estimated 5.37% more viewers compared with baseline predictions

During data wrangling single events were removed such as the Democratic National Convention

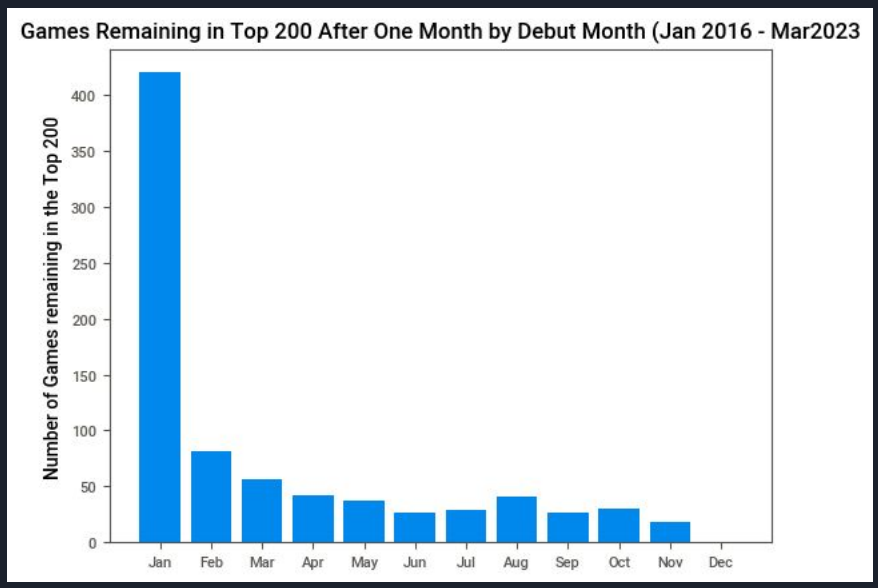
The Republican National Convention

And gaming conventions like Twitchcon

Multiple versions of games were listed separately in the dataset. These have small differences, but the differences may matter to their audience. Thus, these were not combined. However there were instances of actual duplicates which were combined



Approximately a third of the games in the dataset first appeared in the top 200 in January. Granted the dataset starts in January of 2016; however, even subtracting the 200 from the January total it would still be more than double February, it's nearest competitor. And games debuting in the top 200 in January make up approximately 1/3 of the games in the top 200 from Jan 2016 - Mar 2023



Games debuting in January out performed games debuting in all other months. They account for ~52% of the games that remained in the top 200 in the next month. Another surprise is not not a single game that debuted in the top 200 in December remained in the top 200 next month. January debut month was added as a feature to the machine learning model.

Baseline Prediction for Top 200 Prediction

- 809 games continued to be in the top 200 1 month after their debut
- The top 809 games ranked by hours watched were predicted to remain in the top 200, and the rest were predicted to drop from the top 200
- A similar analysis was performed for 3 and 6 months after debut

Machine Learning Model for Top 200 Prediction

- Logistic Regression and Random Forest Classifier were both assessed for 1 month after debut
- Random Forest was slightly better, but not significantly. It was used for the remaining analyses

'max_depth': 7, 'min_samples_leaf': 5,

Accuracy of Baseline and Model 1, 3, and 6 Month After Debut in the Top 200

	1 Month Accuracy	3 Month Accuracy	6 Month Accuracy
Baseline Prediction	68%	83%	100%
Random Forest	80%	90%	100%

Data from 1 month after debut was used to predict whether a game would be in the top 200 3 months after debut and likewise the 3 month data was used for 6 months after debut. Hence the increase of accuracy all the way up to 100% at 6 months after debut.

Baseline and Random Forest Regression

- A regression analysis was performed to attempt to predict hours watched one month after debut for games that remained in the top 200 after one month
- A Random Forest Regression was compared to linear regression for the 1 month data
- The Random Forest Regressor greatly out performed the linear regression

Baseline And Random Forest Regression

- Random Forest Regressor was used for the remaining analyses
- Two baseline predictions were used to compare the model
  1) Hours watched during the debut month adjusted by the median difference of hours watched between the time points
  2) Hours watched during debut month

1 Month After Debut Regression Results

	Median Based Prediction	Hours Watched Prediction	Random Forest Regressor
r^2	.37	.37	.57
MAPE	115%	121%	105%

3 Month After Debut Regression Results

	Median Based Prediction	Hours Watched Prediction	Random Forest Regressor
r^2	.56	.50	.71
MAPE	149%	122%	106%

6 Month After Debut Regression Results

	Median Based Prediction	Hours Watched Prediction	Random Forest Regressor
r^2	.48	.90	.58
MAPE	122%	55%	65%

As you can see the Random Forest Regressor performs modestly better than the baseline prediction up until the 6 month point, where a game that has been popular for 3 months is virtually guaranteed to be popular for another 3 months. Only data from 1 month and 3 month was used for the 6 month predictions

Conclusion

- The top 200 games from Jan 2016 - Mar 2023 have accounted for approximately 91% of Twitch viewership
- Our Classification Model was 12 percent more accurate than the baseline prediction and
- The increase in accuracy likely comes from the games ranked between 136-160 each month

In 2022, the last calendar year 25% of the games that broke into the top 200 did so in January, 75% remained popular in Feb.

Conclusion

- These games account for approximately 5.37% of Twitch viewership, increasing our ad exposure accordingly
- Finally each year in January, we should assess popular games in near real time to get an early lead on the popular games for the next several months

In 2022, the last calendar year 25% of the games that broke into the top 200 did so in January, 75% remained popular in Feb. For 2023 until March, half the games that broke into the top 200 did so in March and 59% remained in the top 200.