Predicting Popularity of Games Streamed on Twitch

Daniel Ansted

Executive Summary:

The top 200 watched games on Twitch account for ~91% of the total hours watched on Twitch. Our machine learning model performed approximately 12 percentage points better than baseline predictions
Thus, we estimate the campaign strategy would target 5.37% more viewers.

Introduction:

Other companies have capitalized on twitch ads. For instance, “Never Stop Gaming” an integrated campaign of Wendy’s and UberEats increased total orders by 9% during the 5 day promotional period and a 15% incremental increase to new UberEats users to Wendy’s in 2020.[1]

Wendy’s and Uber’s successful campaign focused on five popular streamers, and was an intense marketing campaign with a large investment in time and money. In addition Wendy’s followed this Twitch campaign by creating their own Twitch channel. The advertisement campaign under consideration would be more modest and target specific games and/or select upcoming streamers.

The source of the data comes from a Twitch dataset on Kaggle.[2] It contains the hours watched, average viewers, average channels, and other stats of the top 200 streamed games from Jan 2016 - Mar 2023. The primary goal is to predict which games will continue to be in the top 200 games streamed after their debut in 1, 3, and 6 month intervals. The secondary goal is to predict the hours watched assuming they stay in the top 200. The primary goal was analyzed by Random Forest Classifier and the secondary goal using Random Forest Regressor.

Data Wrangling/Cleaning:

There were several single events that broke into the top 200 for hours watched on Twitch, e.g the Republican and Democratic National Conventions, E3, and Twitchcon. These were removed from the dataset as they would not be streamed in the following month, potentially adding unnecessary noise.

There were a few games with special characters that did not appear correctly and sometimes inconsistently. For instance, the various Pokémon games had various incorrect versions of the special character ‘é’ that is in its traditional spelling. This and other similar instances were fixed before data was analyzed.

There were also many instances of virtually identical games being streamed. For instance, Legend of Zelda: Ocarina of Time had three versions that variously broke into the top 200

---

1 https://adage.com/creativity/work/never-stop-gaming/2341476#:~:text=The%20integrated%20campaign%20and%20streaming,minutes%20watched%2C%20and%20343%2C462%2C898%20impressions.

2 https://www.kaggle.com/datasets/rankirsh/evolution-of-top-games-on-twitch

games streamed. The decision for the vast majority of these were to leave as such. The one exception was Final Fantasy XIV.

Final Fantasy XIV was a special case because insofar as an MMORPG can be a single game over time, it was. Games of the same genre, such as World of Warcraft did not have similar issues. There was one instance where two versions of this name were in the top 200 at the same time. Even outside Final Fantasy XIV there were 10 games that had duplicates in a given month. The combinations of these data had data implications for the data for all games.

For duplicate games values such as hours watched and hours streamed were simply combined. Other values used the conservative strategy of choosing between the higher value. This included: peak viewers, streamers, and peak channels. For consistency average viewers and average channels were replaced by new calculations.

One final metric, average viewer ratio, was amended due to sometimes being divided by 0. In this case I simply took the inverse as well as replacing the original data with this calculation for all rows. An additional reason for this decision was that the data card did not supply sufficient provenance for these data points.

To get the data to answer the primary question. A new dataframe was created that primarily consisted of the data for the game's debut month into the top 200 and hours watched from 1 month, 3 months, and 6 months after the debut month. If a game was not still in the top 200 the value would be 0, otherwise the value would be the hours watched for the month in question. For the purposes of classification any value above 0 would be considered as staying in the top 200.

Exploratory Data Analysis

The python library Sweetviz was used to greatly simplify data visualization. Similar to a heatmap it provides correlations on all features with each other. One surprising finding in the exploratory data analysis was that January was the most popular month for a game to make its debut in the top 200 and date was weakly correlated with hours watched in the debut month as well as 1 month, 3 months, and 6 months after debut. Whether or not a game debuted in January was used as a feature in the models as this seemed to be the driving factor for the impact of date.

Preprocessing

Before modeling, it was necessary to remove any games that debuted in the last month of the dataset as these would automatically be classified as being dropped out of the top 200 and this would often be erroneous. Similarly games that debuted after December 2022 were removed from the 3 month classification and regression tasks and September 2022 for the 6 month tasks.

Of note the regression was only attempting to predict the hours watched in these future months for games that remained consistently in the top 200 for the months in question. For both regression and classification a standard train_test_split from scikitlearn was implemented. No scaling was done for either the Random Forest Classifier or Random Forest Regression as these are robust to outliers and non-normal distributions.

Modeling:

A naive baseline model was developed to compare against the Random Forest Classifier. First using historical data it was determined that approximately 40% of games remained in the top 200 in the next month. Second, the games were ordered by hours watched in descending order. Finally the top 40% of games in terms of hours watched were predicted to remain in the top 200 in the next month, the rest were predicted to drop. This naive baseline model had an accuracy of 68%. A similar procedure was run for the 3 month and 6 month time points.

A Random Forest Classifier was chosen as the machine learning model and grid search was used for hypertuning the parameters. And results against the baseline are in the chart below.

Accuracy	Baseline (based on hours watched)	Random Forest Classifier
1 Month after Debut	68%	80%
3 Month after Debut	83%	93%
6 Month after Debut	100%	100%

The goal of the regression analysis was to predict the hours watched for all games that continued to be popular after the 1,3,6 month intervals. Models performed slightly better than baseline predictions, with the exception of the 6 month interval. There were two baseline predictions that were used for each regression analysis. First, I found the difference between the hours watched one month post debut and the debut month, then added the median of this difference to the hours watched to get a prediction for one month post debut month. The second baseline used the hours watched from debut month as the prediction for hours watched one month after debut.

The model used a Random Forest Regressor and hypertuning the parameters did not produce better results, thus out of the box Random Forest Regressor was compared with these baseline predictions. Another attempt to produce a better model was to attempt to predict the hours watched/total hours watched on Twitch. This resulted in only slightly better r^2 scores so this approach was not pursued further. The results for the predictions and model are in table 1, mean absolute percentage error and r^2 were chosen as the primary metrics.

Table 1

1 Month After Debut Results:

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

3 Months After Debut Results:

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

6 Months After Debut Results:

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

Conclusion:

The Random Forest Classifier predicts which games will remain in the top 200 after 1 month 12 percentage points better than the baseline prediction and 10% more accurate than baseline prediction three months after debut. However, given that the three month baseline prediction is much more accurate this has little to no business impact.

The model predicting which games will be in the top 200 after 1 month may have a modest but important business impact as it also influences predictions for if a game will continue to be in the top 200. A similar modest improvement over the baseline prediction was found using a Random Forest Regressor for the 1 and 3 month endpoints.

Another salient feature that should be noted is the importance of the January debut month. 85% of the games that remained in the top 200 after 1 month debuted in January. This is a surprising finding in the data and suggests that we should consider this timing when we are considering marketing strategies.

Similar, but in a more modest fashion to the Wendy’s case study, we might find some mid-tier streamers who have shown consistency in early adoption of games into the top 200 and sponsor them.

Appendix 1:

Figure: 1 Confusion Matrix for Random Forest Classifier at 40% threshold

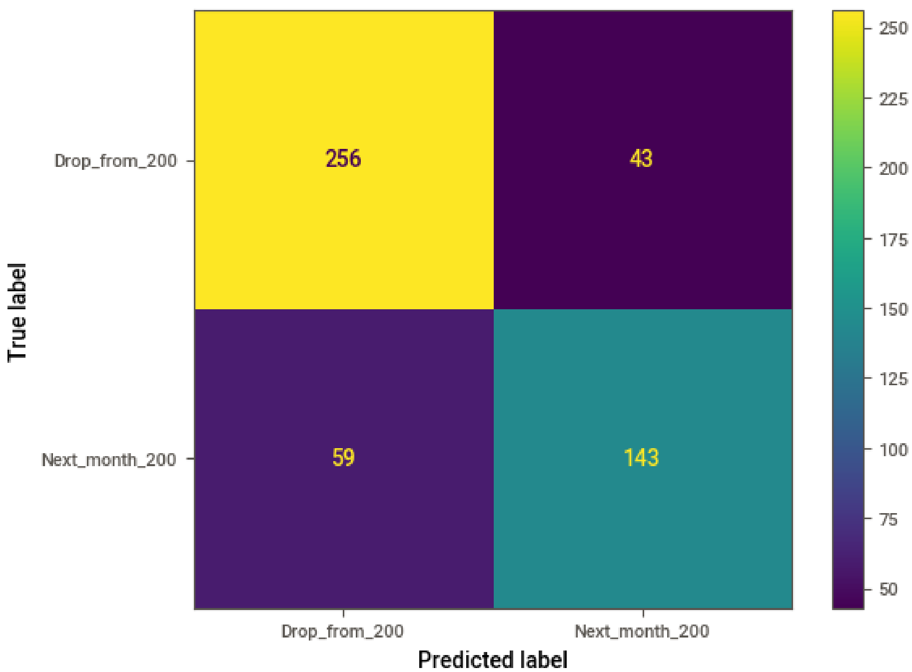
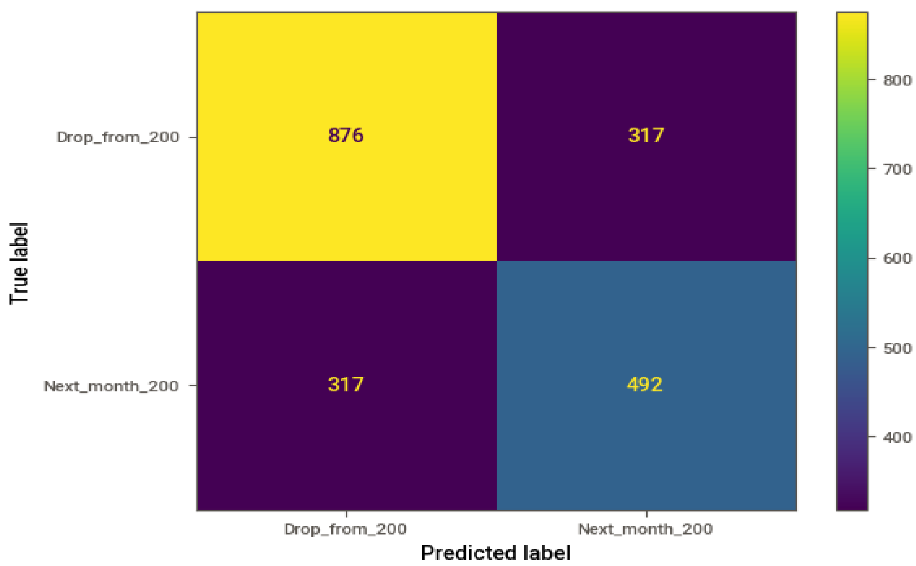


Figure 2: Confusion Matrix for baseline prediction



True label
Drop_from_200
Next_month_200
876
317
317
492
Drop_from_200
Next_month_200
Predicted label
800
700
600
500
400