

Analysis of Creator Focused Media Websites

Owner: Daniel Cavazos

Date: 06/22/2022

Table of Contents

EXECUTIVE SUMMARY	2
WHO AND WHAT GETS VIEWS?	2
RECOMMENDATIONS	4
BACKGROUND	5
CREATOR FOCUSED MEDIA WEBSITES.....	5
ANALYSIS.....	5
TWITCH.....	5
Overview and Small Creators.....	5
Paid Promotions	7
Game Categories.....	8
YOUTUBE	9
Overview and Smaller Channels	9
Music On YouTube.....	10
Trending Categories	10
NEXT STEPS.....	11
APPENDIX	13
LIMITATIONS	13
Unclean Twitch Data.....	13
Twitch Data Collection Issues.....	13
Category Names	14
Twitch Drop Analysis	14
DEFINITIONS	14
FURTHER RESEARCH OPPORTUNITIES	15
Click Bait	15
Twitch Metas / Trends Over Time	15
THE DATASET	15

Executive Summary

For most creator focused media websites¹, revenue is earned by showing viewers advertisements; therefore, increasing viewership increases revenue. To analyze who and what increases viewership, we looked towards two titans of the industry, Twitch and YouTube. Twitch is a livestreaming service owned by Amazon which focuses on video game media content. YouTube is a video sharing platform specializing in pre-recorded videos owned by Google. Contrary to Twitch, YouTube does not have a specific focus like video games and hosts a wide variety of content.

Who and What Gets Views?

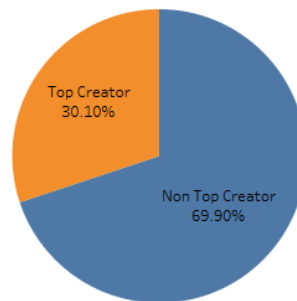
The YouTube and Twitch data we chose to look at looked at around 41,000 trending YouTube videos and 74,000 Twitch streams over their respective time periods. Analyzing these distributions, we found that both were right-skewed, meaning that the majority of views were in the lower end of the distribution, however, the top channels' and streamers' view counts heavily influenced the average views per media occurrence. Looking at the median values for these services provides a better understanding of the center of our distributions.

Service	Time Period	Number of Media Occurrences	Mean	Median
YouTube	7 months	40,949	2,360,785	681,861
Twitch	About 2 months	73,909	1,499	355

Table 1: Summary Statistics

Top creators heavily influence the distribution of views on creator focused social media websites. For both Twitch and YouTube, the top 1% of creators were responsible for about 30% of total viewership across their respective platforms. In other words, 15 Twitch streamers and 22 YouTube channels were responsible for three out of every ten views for Twitch and YouTube.

YouTube Views By Creator Status



Twitch Views By Creator Status

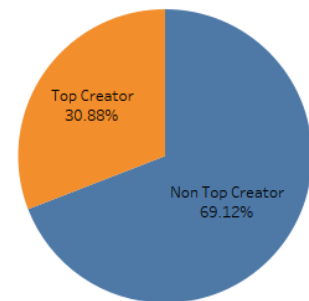


Figure 1: Views By Creator Status

¹ Find the definition of content focused media websites in the [Definitions](#) section of the Appendix.

Looking more in-depth at these 22 YouTube channels, very few are creator-owned channels with a substantial portion being channels owned by the company VEVO. These VEVO-owned channels² account for about 25% of total trending views meaning one out of every four views of a trending YouTube video were on a VEVO-owned channel. The top streamers and channels are great sources for high view counts; however, their dominance can overshadow the smaller creators. One potential method to increase discoverability³ implemented by [Twitch are Drops](#), a reward viewers receive by watching select streams. Unfortunately, the majority of streamers found little to no increase in their average viewership when promoting Drops on their stream.

One method of distributing viewership both services utilize is Categorization. Twitch groups views by the game the streamer is playing and YouTube groups views by predefined categories. There, however, does not seem to be a correlation between the popularity of these categories⁴ and the average views present. This insight then insinuates there is more to deciding what to watch than simply the category the media belongs to. These categories failing to increase discoverability then hurts smaller creators.

To increase the viewership over an entire platform, a focus should be placed on smaller creators. There are significantly more creators with smaller view counts. Attempting to increase the average viewership of top creators would only see marginal increases. Alternatively, attempting to expose more viewers to smaller creators would increase the average viewer community of that creator, making another reason for viewers to return to the website. Having multiple regularly viewed creators should keep users returning to the site on a more regular basis and therefore viewing more advertisements.

² Find the definition of VEVO versus NON VEVO channels in the [Definitions](#) section of the Appendix.

³ Find the definition of discoverability in the [Definitions](#) section of the Appendix.

⁴ Find the definition of popularity in the [Definitions](#) section of the Appendix.

Recommendations

Based off of the insights gained from our analysis, there are four actions we recommend to further our understanding of key factors related to viewership on creator focused media websites.

Insight	Recommendation
The top 1% of creators hold a majority of the total viewership on both platforms.	Investigate methods of increasing discoverability for smaller creators. An in-depth analysis on Twitch's hosting and raiding systems could explore this subject.
Larger companies can overshadow creators.	Test if creating separate trending pages for company-owned channels versus creator-owned channels increases overall viewership.
Category popularity and game popularity do not seem to represent the average views of each section.	Research alternative methods of discoverability besides categories to support smaller creators on the platform.
Promotion from companies (Twitch Drops) have largely no impact on creator's viewership.	Investigate further with more applicable data and a more in-depth analysis to adapt future promotion approaches.

Table 2: Recommendations

Background

Creator Focused Media Websites

Media websites are websites designed to promote sharing media with other people. Creator focused media websites specifically promote interactions between the person sharing the media (the creator) and those consuming the content (the viewer). The two main creator focused media websites are Twitch and YouTube. Twitch is a live streaming service owned by Amazon that focuses on broadcasting game livestreams, specifically video game livestreams. While traditional games such as chess and non-game related content categories (e.g., Just Chatting) find success on Twitch, the majority and focus of the content is centered around video games. YouTube is a video sharing website owned by Google. Unlike Twitch, YouTube does not have a specific focus like video games and has trending videos in various categories such as music, sports, politics, and gaming. Additionally, YouTube specializes in pre-recorded videos rather than livestreams. While YouTube does have a livestreaming service on top of their normal content, we will not be analyzing this service in this report.

Analysis

The titans of the creator focused media websites, Twitch and YouTube, have found success due to their ability to increase overall viewership. The focus of this analysis will be to find areas of these services that have an impact on viewership and to use these trends to fuel further investigation on adapting and evolving creator focused media website features.

Twitch

Overview and Small Creators

For the Twitch data, there were 3,036,497 observations and six features after cleaning the dataset. In these observations, there were 73,909 unique livestreams. These livestreams spanned across about two months, from September 22, 2020, to November 6, 2020. Through these two months, there was a minimum of zero views to a stream and a maximum of around 385,000 views. In this wide range of data, there was a mean of around 1500 views per stream and a median of 355 views per stream. This difference between the averages suggests right-skewed distribution. We can see this right-skew in figure 1 and figure 2 (note: jitter has been added to figure 1 due to the number of observations). In these visualizations, we can see that the distribution of this data is very wide. Despite the width of the data, the vast majority of the stream views exist towards the beginning of the distribution. The second visualization displays this showing that 95% of the stream views are below 5,155 views. This visualization also shows the hard right-skew of the distribution to greater effect here. These figures demonstrate that the

majority of streams and creators on Twitch stream to smaller audiences and that it is exceedingly rare to have streams with more than 5,155 people viewing at a time.

Twitch Stream Views Distribution

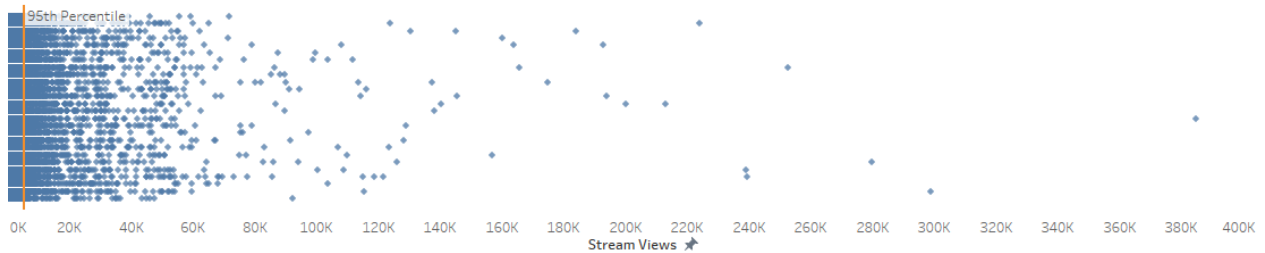


Figure 2: Twitch Stream Views Distribution

Twitch Stream Views Distribution
Histogram of stream views below the 95th Percentile

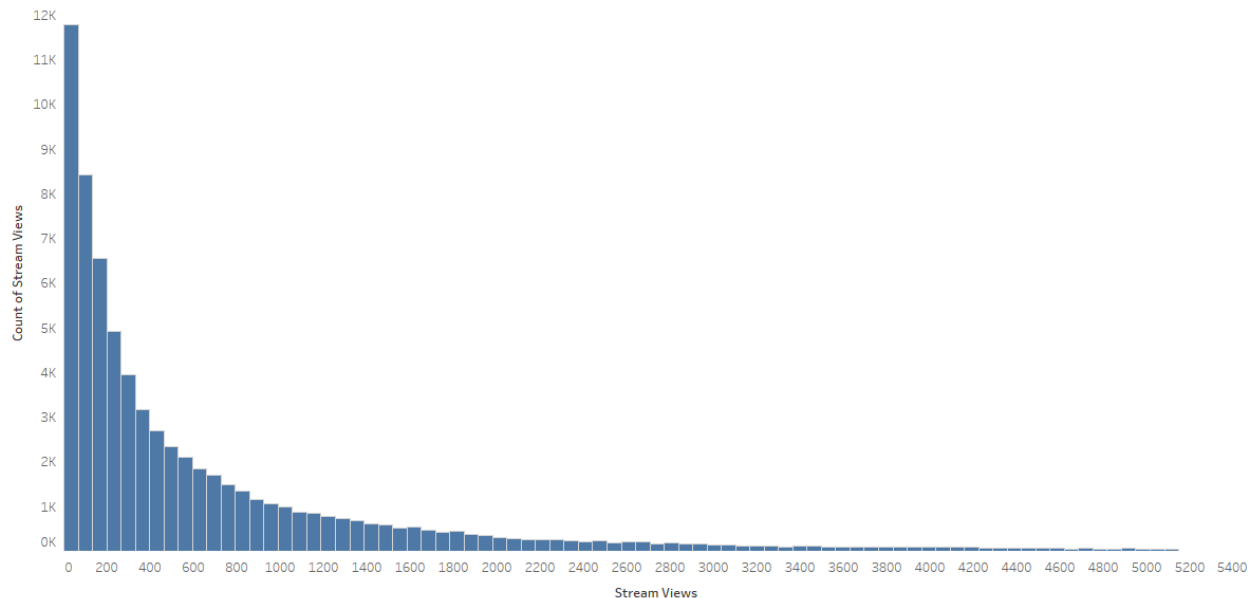


Figure 3: Twitch Stream Views Below the 95 Percentile Distribution

Despite the majority of streams and creators having fewer views, the majority of views across the data are generated by the top 1% streamers⁵. In the data, the top 1% of streamers, consisting of fifteen unique creators, were responsible for 31% of the views across all streams. In other words, fifteen of the 1,516 streamers were responsible for about three out of every ten views on Twitch, showing smaller streamers can be overlooked by the general viewer population.

⁵ Find the definition of top creators / 1% of creators under the [Definitions](#) section in the Appendix.

Paid Promotions

One feature available to Twitch streams is the ability to add Drops to their stream. These Drops are rewards given to viewers if they watch the stream for long enough and can be enabled so long as the streamer is playing or interacting with the game that sponsored the Drops. Normally video game companies sponsor these Drops when there is a major in-game event or esports tournament happening. For example, when the video game Rocket League updated and released a new time-limited game mode, they sponsored Drops which streamers who were playing Rocket League could enable.

Twitch streamers enabling Drops should then theoretically improve viewership as it encourages users to visit and watch these streams. To assess this hypothesis, we parsed the titles of all streams in our data to look for mentions of Drops and compared the average views for streamers with and without Drops enabled. Figure 3 shows the distribution of these differences where a positive number represents an increase in viewership when Drops were enabled and a negative number representing a decrease in viewership. About 80% of streamers who at some point mentioned Drops in their title saw little to no change in their average viewership. This brief analysis insinuates Drops as paid promotions do little to increase viewership for the majority of streamers⁶.

Drops Stream Views and Standard Stream Views Difference Distribution
Histogram of Average Stream Views With Drops and Average Standard Stream Views Difference

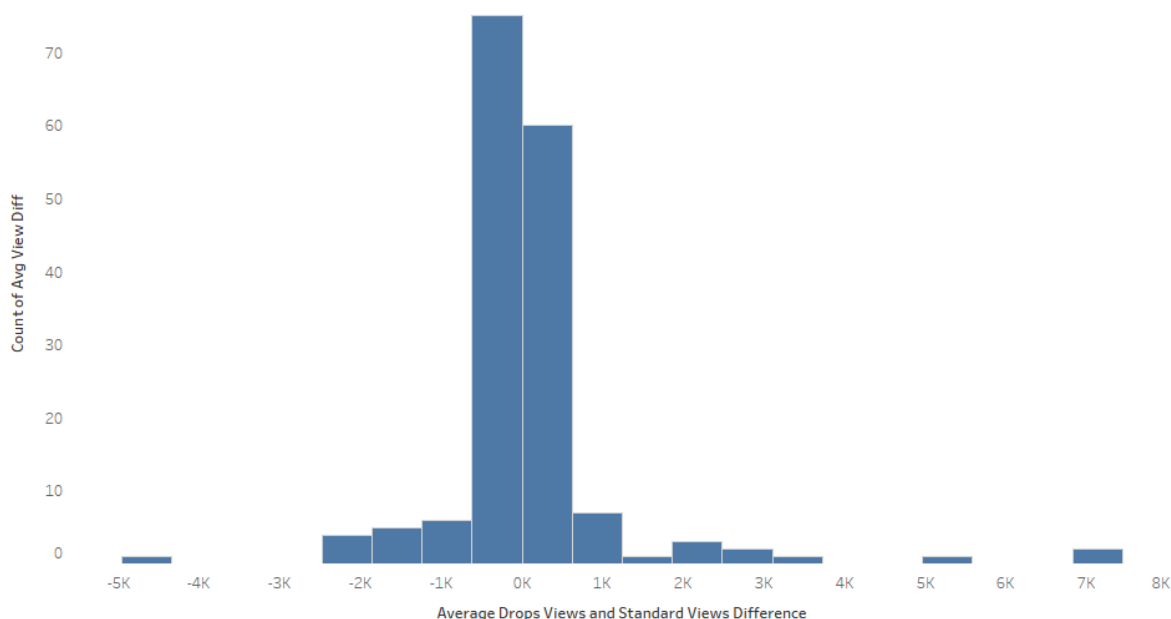


Figure 4: Drops Stream Views and Standard Stream Views Difference Distribution

⁶ For caveats and limitations to this analysis visit the "[Twitch Drop Analysis](#)" section in the Appendix.

Game Categories

There are 954 unique games present in our dataset. To gain insight into relationships between these games and viewership, we looked at the top twenty-five most popular games on Twitch⁷. Figure 4 shows popularity does not correlate with the average views a stream receives. It also shows several extreme outliers such as Call of Duty: Modern Warfare with the highest number of streams but a small number of average views at about 1,163 and Minecraft with the 13th highest number of streams and a large number of average views at about 4,150 views per stream. This metric displays there is more to what viewers want to watch than just the game a streamer is playing and that identifying new ways to introduce discoverability besides game categories, particularly for smaller streamers, is important.

Views For Top 25 Most Popular Games

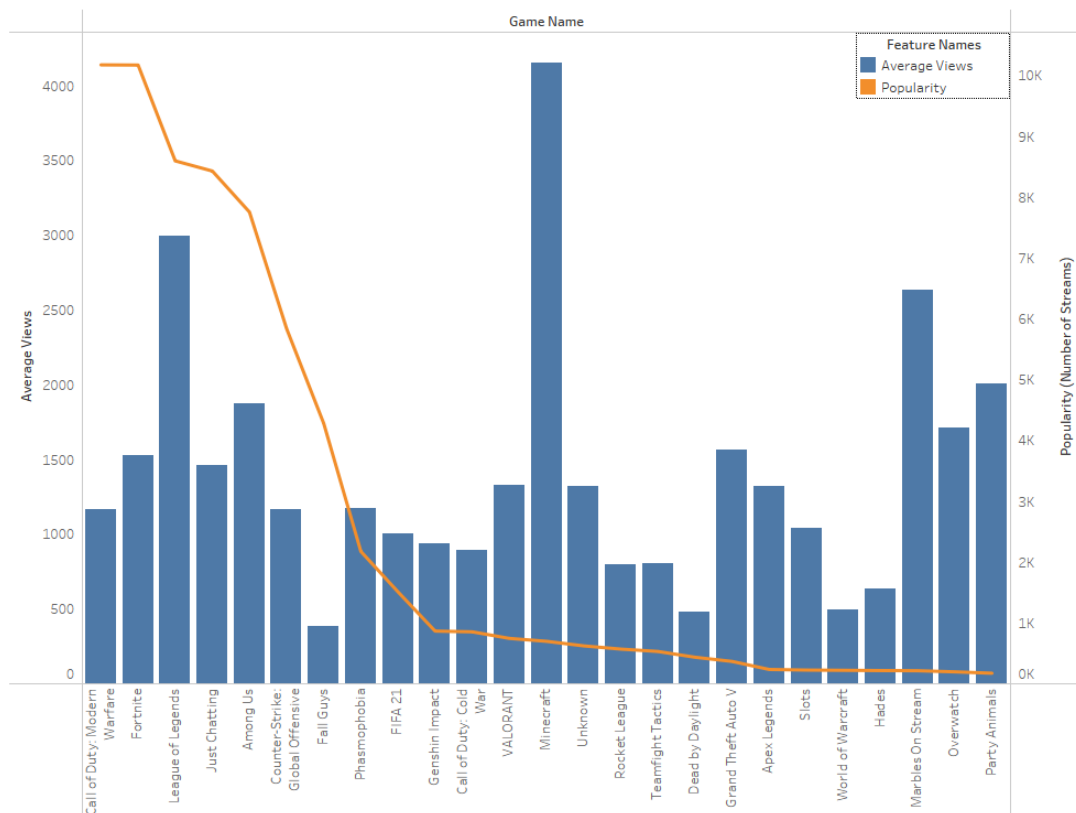


Figure 5: Views For Top 25 Most Popular Games

⁷ Find the definition of popularity in the [Definitions](#) section of the Appendix.

YouTube

Overview and Smaller Channels

The YouTube data consisted of 40,949 observations and thirteen features with each observation representing a trending YouTube video. Data was collected for 7 months ranging from November 14, 2017, and June 14, 2018. In the data, the lowest views for a trending video was 549 views and the greatest number of views for a single video was 225,211,923 ([Childish Gambino's "This is America" music video](#), now under the channel name Donald Glover). The mean number of views for a trending video was around 2,360,785 with median views being 681,861. Similar to the Twitch data, the wide range and significantly larger mean than median suggests a right-skewed distribution shown by Figures 5 and 6. Figure 6 shows in more depth the distribution of data below the 95th percentile, 9,017,287. Much like the Twitch data distribution, the majority of trending videos are found below this nine million views line.

YouTube Trending Video Views Distribution

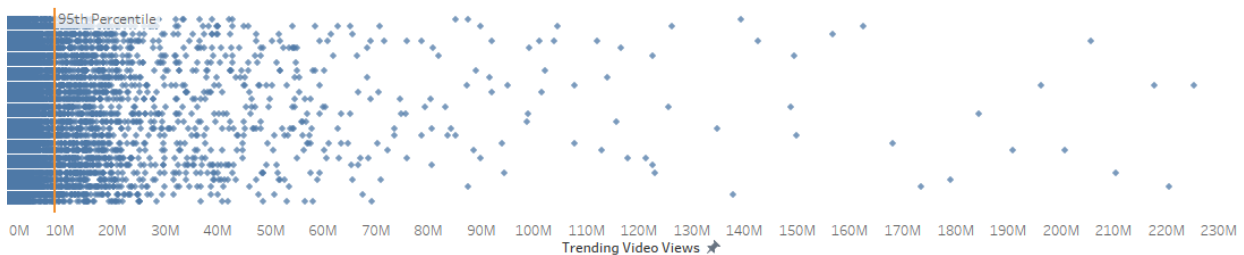


Figure 6: YouTube Trending Video Views Distribution

YouTube Trending Video Views Distribution
Histogram of trending video views below the 95th Percentile

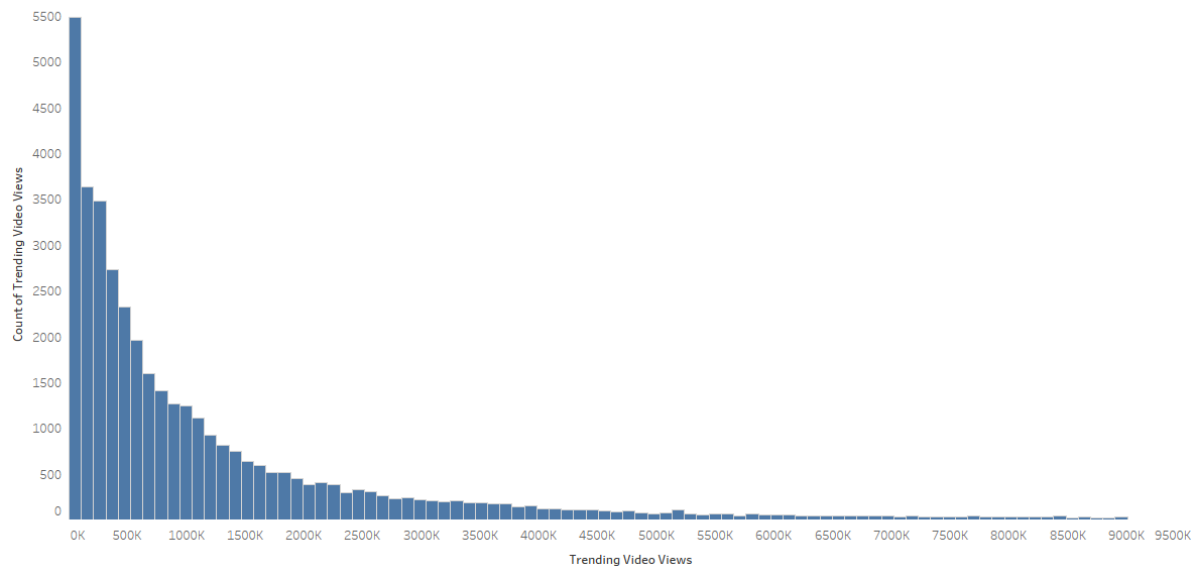


Figure 7: YouTube Trending Video Views Below the 95 Percentile Distribution

Regardless of the majority of creator's trending videos existing below nine million views, the top 1% of channels are responsible for a substantial portion of the total views. Out of the 2,207 unique channels in the dataset, twenty-two channels hold about 30% of the total trending video views. This shows how the top YouTube channels can overshadow smaller channels from the general viewer population.

Music On YouTube

Despite YouTube having a reputation in the past for supporting smaller channels, a sizable portion of YouTube and YouTube's trending videos consists of business channels. While some creators exist in the 22 channels holding 30% of the total trending video views, the majority are company channels such as [Marvel Entertainment](#) or [Universal Pictures](#) and large music stars such as [Maluma](#) or [Ariana Grande](#). The company [VEVO](#), a joint venture between major record companies, started with the underlying idea of hosting music videos for their musicians on YouTube. Several of the top channels were channels owned by VEVO for various musicians such as ZaynVEVO, ArianaGrandeVEVO, or TheWeekndVEVO. Looking through the entire dataset, channels with the word VEVO in them⁸ accounted for 26.30% of the total views as seen on Figure 7. For many people, YouTube is a website to watch these music videos, however, these company-owned channels can stifle smaller channels as their dominance in trending videos can overshadow these creators. As music videos occupy more trending videos spots, less may be claimed by creator-owned channels.

Percentage of Views From VEVO Channels

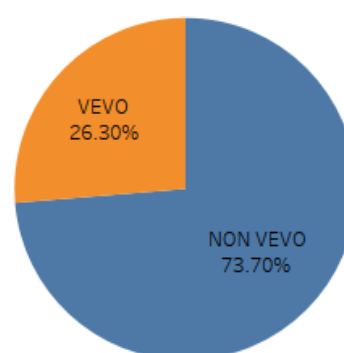


Figure 1: Percentage of Views From VEVO Channels

Trending Categories

One way YouTube attempts to diversify their trending page is by categorizing the trending videos. There were sixteen unique categories in the United States' trending page during the given time period. Looking at the relationship between popularity and average views in Figure 8, there does not appear to be a correlation between the popularity of this category and average views. This insinuates there is more to what users choose to view than popularity of a category and that it may be beneficial to separate these categories into their own trending pages or inspect whether they are useful categories to define.

⁸ Find the definition of VEVO versus NON VEVO channel in the [Definitions](#) section of the Appendix.

Average Trending Video Views By Category

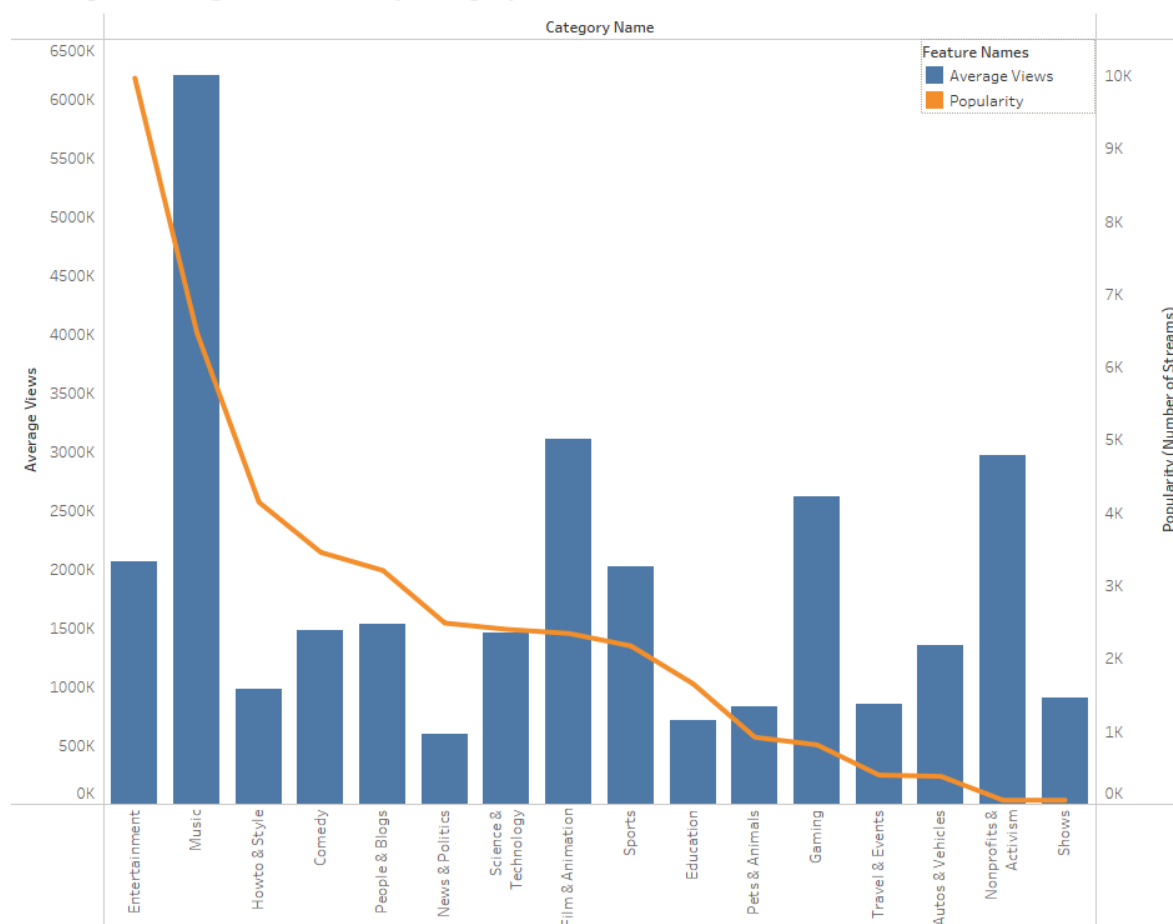


Figure 8: Trending Video Views By Category

Next Steps

Recommendation: Investigate methods of increasing discoverability for smaller creators.

The top 1% of creators hold a significant percentage of total viewership, around 30%, on both YouTube and Twitch. Additionally, when looking at the average viewership per category on YouTube and the average viewership per game on Twitch, there appeared to be no correlation between the popularity of the category or game and the average viewership. Both of these insights harm discoverability for smaller creators, therefore, investigating alternative methods in increasing discoverability should be promoted. One potential avenue is an in-depth analysis on Twitch's hosting and raiding systems. With this analysis, we

could find if these systems aid in viewers finding smaller creators and overall increasing viewership.

Recommendation: Test if creating separate trending pages for company-owned channels versus creator-owned channels increases overall viewership.

Analyzing YouTube's trending videos, we found that larger company-owned channels can tend to overshadow creator-owned channels. Further analysis should be done testing if there is a significant difference in overall average viewership and average viewership within categories for a newly designed trending page splitting channels by company-owned and creator-owned.

Recommendation: Investigate Twitch Drops with more applicable data and a more in-depth analysis to adapt future promotion approaches.

When analyzing the Twitch's Drops promotion method, we found that the majority of streamers advertising they had Drops enabled saw little to no increase in average viewership. Drops are a potential method for streamers to increase their viewership and attract new viewers, however, due to limitations with the dataset, we were not able to confidently determine whether there was a change in average viewership. With a more robust dataset it could be possible to perform statistical tests to more accurately understand if there is a difference between average viewership when a streamer had Drops enabled versus when a streamer did not have Drops enabled.

Increasing overall viewership is the primary goal for creator focused media websites. While top creators hold a substantial portion of total views, focusing efforts on increasing the average viewer base of smaller content creators would be better for the overall average viewership of the service.

Appendix

Limitations

Unclean Twitch Data

The application used to read in comma delimited data did not process the data properly. This resulted in an incomplete dataset, about 5% of the data was lost. While this is a sizable portion of the dataset, we are content with using the remaining 95% of the data as this consists of 3,036,497 unique observations. To clean the data, we used the two features formatted as dates (TimestampUTC and StarttimeUTC) as gatekeepers. When the processor incorrectly delimited the data, these features were often filled with non-date like values. Therefore, if both of the values of these features were formatted properly in a single observation it remained and if one or both of these were improperly formatted it was removed from the table. For the code used to create the cleaned data see figure 9.

```
-- create a new table with the clean data
DROP TABLE clean_twitch_data;
CREATE TABLE clean_twitch_data AS
SELECT
    Name,
    TimestampUTC,
    CurrentGameID,
    StreamTitle,
    ViewerCount,
    StarttimeUTC
FROM twitch_stream_data
WHERE
    (
        STRFTIME('%Y-%m-%d %H:%M:%S',TimestampUTC) IS NOT NULL
        AND
        STRFTIME('%Y-%m-%d %H:%M:%S',TimestampUTC) NOT LIKE '-471%'
    )
    AND
    (
        STRFTIME('%Y-%m-%d %H:%M:%S',StarttimeUTC) IS NOT NULL
        AND
        STRFTIME('%Y-%m-%d %H:%M:%S',StarttimeUTC) NOT LIKE '-471%'
    );
```

Figure 9: Clean Twitch Data SQL Code

Twitch Data Collection Issues

Due to the process used to collect the Twitch data⁹, one observation does not represent one stream. We chose to define a new stream as the unique combination of the streamer name, the time they started their stream, and the game they were playing. This means if a streamer changed the game, they were playing mid-stream, we counted it as a completely separate livestream for the purpose of analysis.

⁹ See "[The Dataset](#)" section of Appendix for more information.

Category Names

There was unfortunately no data present defining a relationship between CurrentGameID and the name of the game this identifier represents. To briefly overcome this limitation, we searched through the titles and visited some Twitch VODs to find the related game name. These names were then hard coded using the SQL CASE statement. This process was used only for the top twenty-five most popular games. To expand this process to all of the games present in the dataset, the Twitch API would need to be utilized to create a table with a relationship between CurrentGameID and the GameName.

Similar to the Twitch table, there was no data originally present in the dataset for the YouTube category names. Unlike the Twitch data, the original collector of the YouTube data collected this information in a [separate file](#). Referencing this file, we were able to implement the category names for analysis.

Twitch Drop Analysis

Due to not having data on whether a stream had Drops enabled or not, we used the streamer's title to estimate this. If a streamer had the text "drop" (non-case sensitive) in their title, we considered that stream to have Drops enabled. Due to this limitation, we only had 168 unique streamers (out of 1,516 total unique streamers) who had at some point enabled Drops on their stream. We are cautious about the conclusions we made during this analysis and promote a further investigation with more in-depth analysis and features.

Definitions

Creator focused media website: Media websites are websites designed to promote sharing media with other people. Creator focused media websites specifically promote interactions between the person sharing the media (the creator) and those consuming the content (the viewer).

(Content) Creator: A user of a creator focused media website who shares media to viewers.

Discoverability: We define discoverability as how easy it is for a new viewer to find a content creator. This is **not** a defined measurable metric and is only a term used to define a concept.

Top creator / 1% of creators: We chose to represent top creators and the 1% of creators as creators whose total views fall within the top 1% of total views across their respective platform.

VEVO vs. NON VEVO channels: We defined VEVO-owned channels as any YouTube channel with the text "VEVO" (non-case sensitive) in it.

Drop Sponsored Stream: We defined a drop sponsored stream as one which had the text “drop” (non-case sensitive) in the title.

Popular Category / Game: We defined a popular category / game as one with a large number of unique pieces of media. For example, a popular game on Twitch would have a lot of streams with it listed as the game being played.

Further Research Opportunities

Click Bait

Click bait is a term used to describe when a content creator purposefully overexaggerates in their video thumbnail or title to increase the odds a viewer clicks on the video. An interesting point of research could be to look at how viewership changes based on the ratio of uppercase letters to lowercase letters in the title. Expanding on this, one could also look into common phrases used in click bait and see if viewership changes if these phrases are used in media titles.

Twitch Metas / Trends Over Time

A colloquial term used in the Twitch community is the “Twitch Meta.” This refers to the current popular trend on Twitch, for instance if a new game is exceedingly popular and the vast majority of streamers are playing it, the community may deem that game the new Twitch Meta. It would be interesting to see if these metas could be predicted or if a similar concept can be found on YouTube.

The Dataset

The data for the YouTube trending videos and Twitch streams comes from a dataset on streaming services analysis. The YouTube trending data was collected by GitHub user [Mitchell Jolly](#) using YouTube's API service and shared via [Kaggle](#). In this analysis, we specifically look at the United States trending videos. The Twitch streaming data comes from [Data & Sons](#) and features around 2,500 popular Twitch streamers with real time data dated every 5 minutes. Unfortunately, the original collectors did not provide documentation, therefore we are unsure how they chose to define popular Twitch streamers. Additionally, we are unsure on their collection method but assume the use of Twitch's API service.

An important distinction between our two datasets. The YouTube data consists of **trending** YouTube videos rather than all videos from popular creators. In contrast, the Twitch data consists of **all videos from popular creators** rather than trending or popular streams. Despite this difference, we believe trends within the data are comparable between services, however, we keep this caveat in mind throughout the analysis.