AUTHENTIC MEDIA ASCENSION (AMA)

PREDICTING YOUTUBE VIDEO PERFORMANCE

BY:

MATEO PRICE
JIHAD ESMAIL
LAINEY DOW
KARAN AGRAWAL
LEO SERDAR

NOVEMBER 2020

Table of Contents

Introduction

As a team of analytics and strategy experts that focuses primarily on helping YouTubers grow, we've always wanted to better understand what the early success (or failure) of a video after being posted says about its long-term success. Is a video that starts off poorly bound to stay like that forever? What about a video that starts off hot? Will it remain great, or will things trail off after core subscribers are done watching?

With several of our partners, we noticed that one major difference between different content types for a creator are the traffic sources they pull from. Given that some videos had near-identical metrics, and yet, had way different success levels half a year later, we decided to have a whiteboard session and explore this question more in-depth.

We realized that the best way to answer this question would be to pool together data from multiple creators across different verticals and collect as much data as possible shortly after a video was posted. With this wealth of information, we could run regressions (for those unfamiliar -- they inform how strong a relationship is between two variables, i.e. click-through rate and total views) to find which variables were statistically significant and most-helpful in predicting a video's overall success.

As a result, we set off on a month-long journey to collect this data manually (unfortunately the API doesn't support the information we wanted to gather) and analyze it. Our conclusions mostly supported our initial prediction (more on that below), but we learned a lot in the process and hope this can be of service to you as well.

We hope this information can give you a better sense of how to predict the long-term success of a video early in its lifespan. We're also releasing an Excel model for those who'd like to utilize our findings for their own channels.

Thanks for taking the time to read through. We always love hearing from other creators, so please let us know your thoughts. Happy to dive deeper into this together if people are interested -- DM me in Discord or email me at matthew@ama-digital.com. Cheers!

- ***Mateo Price**, founder at Authentic Media Ascension (AMA)*

Background Information

Problem Statement:

There are many YouTube experts that claim looking at traffic sources early in a video's lifespan can be a secret identifier into whether a video will succeed. This begs the question: do traffic sources in the first couple days after a video is published preview the success the video will have? If they don't, are there any other hidden indicators that do?

Hypothesis:

We hypothesized that there is no serious relationship between the traffic sources on the first day of a video being published and its overall success. We expect that although there may be some relationship closer to the 3-to-7 day mark, the relationship is not statistically significant.

Methodology

To address this question, we analyzed 150 videos across different creators to look for the relationship between what most people define as a successful video (the view count) and independent variables: browse, suggested, notification, search, and channel page traffic sources in the first 24H and first 72H.

The dependent variables will be the total view count a year after the initial video is posted (specifically days 4-365). This will help with confidence in the data quality and ensure that enough time was given to properly see results.

We ran a multi-linear regression to properly analyze the data.

Results

For some creators, there was one traffic sources-related variable shortly after a video was posted that did have a significant impact on long-term video success: the impression count for suggested traffic.

However, this metric, among others, is best-utilized when looked at *in tandem* to the view-count after 72 hours. When looked at alone as a predictor of success, view-count after 72 hours was the strongest predictor of a video being successful.

The Role of Traffic Sources

We started off this white paper with the intention of looking only at the relationship between traffic sources and video performance. We realized early on, however, that it would be most beneficial to look at *all* variables and to see which of them, whether traffic sources or not, had the best ability to predict long-term success.

How Traffic Sources Can Alter a Video's Trajectory

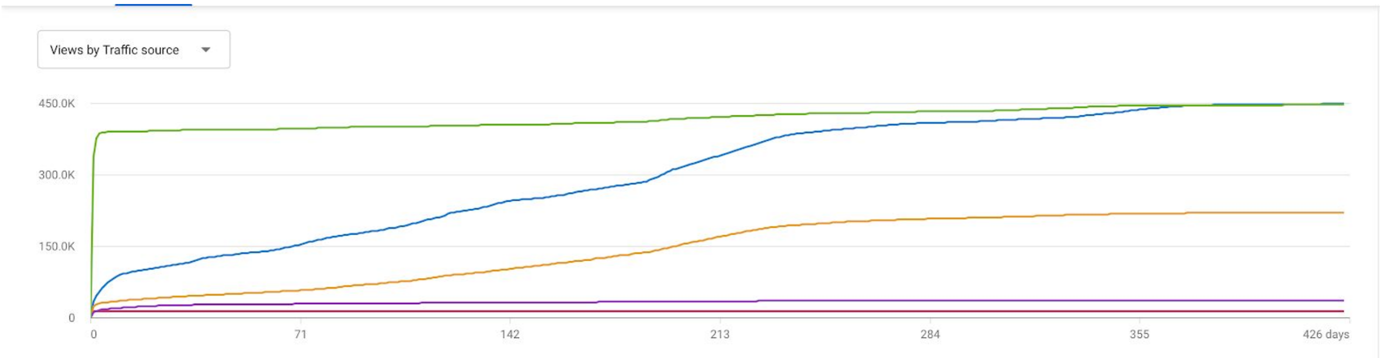
The two strongest traffic sources for most creators are browse features and suggested.

1. **Browse features traffic** refers to viewers coming from the home or sub pages (think of your core subscribers)

2. **Suggested traffic** refers to viewers coming from the 'up next' area of YouTube

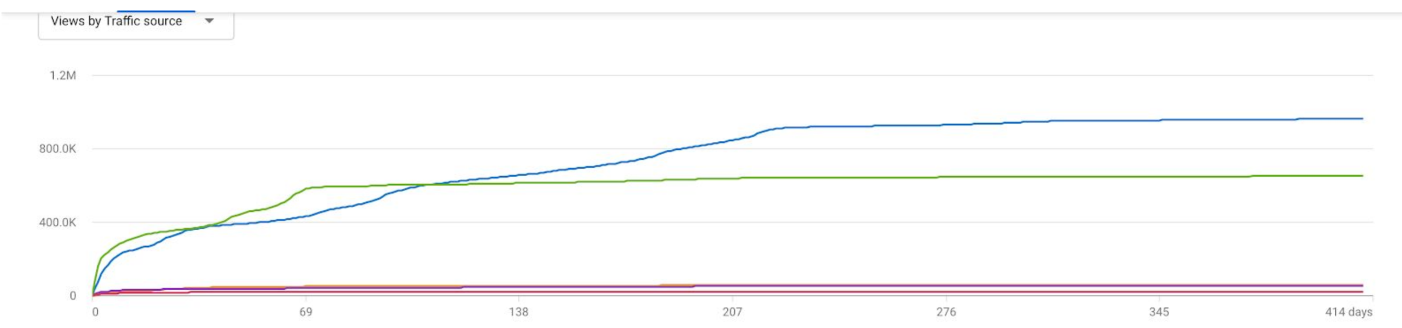
What's interesting is that a given video can have very different traffic source patterns. Some videos reside well with the current audience but no-one else, some videos do the opposite, and some videos do both. When looking at videos that successfully bring in views over long periods of time, you notice some very different graphs than normal.

Examples of a video with long-term success (still bringing in views each day)



Traffic source	Impressions	Impressions click-through rate	Views ↓	Average view duration	Watch time (hours)
Total	**10,495,641**	**7.4%**	**1,204,462**	**5:53**	**118,210.2**
Suggested videos	5,375,471	5.2%	449,307	5:38	42,213.4
Browse features	2,294,832	13.1%	447,545	6:17	46,956.7
YouTube search	2,024,264	8.0%	221,205	5:36	20,664.2
Channel pages	768,476	3.8%	37,343	6:07	3,814.4

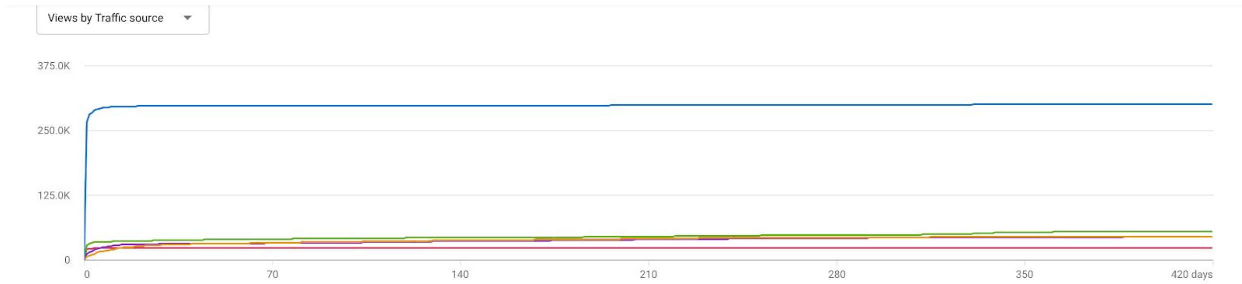
Blue line is suggested; green is browse. Notice: browse plateaus but steady suggested traffic leading to success.



Traffic source	Impressions	Impressions click-through rate	Views ↓	Average view duration	Watch time (hours)
Total	**23,921,577**	**5.0%**	**1,809,130**	**6:07**	**184,461.9**
Suggested videos	17,422,856	3.8%	963,961	6:07	98,422.1
Browse features	5,311,996	9.0%	652,046	6:02	65,696.9
End screens	–	–	58,893	6:39	6,535.5
YouTube search	514,473	7.2%	54,822	5:56	5,423.5
Channel pages	646,863	2.6%	23,032	6:58	2,677.5

Blue line is suggested; green line is browse. Similar story to above.

And below, an example of a video with no short or long-term success



Traffic source	Impressions	Impressions click-through rate	Views	Average view duration	Watch time (hours)
Total	**3,081,380**	**10.2%**	**490,785**	**5:25**	**44,346.8**
Browse features	1,397,820	15.7%	301,373	5:38	28,335.5
Suggested videos	531,462	4.9%	55,052	6:03	5,565.6
Channel pages	729,847	4.8%	45,133	4:20	3,264.7
YouTube search	397,074	8.3%	44,411	4:15	3,147.3
Notifications	–	–	22,868	5:19	2,027.9

The first graph illustrates an example of a video that started off strong and continued over time to bring in a lot of traffic. The second graph shows a slower start that builds up momentum over time. And the third graph shows a strong start that completely dies out.

This leads us back to the question: does knowing early traffic source patterns tell us anything about long-term video success?

Data Collection

Choosing Our Data

When selecting the data to include in our study, we opted in for any early performance metrics that could potentially predict long-term success. The emphasis remained on traffic sources, but we didn't limit our net to that region alone.

The variables we measured:

1. Views after 1 year (the dependent variable)
2. Impressions after 1 year
3. Views 24 hours after posting
4. CTR 24 hours after posting
5. CTR on browse feature traffic 24 hours after posting
6. CTR on suggested traffic 24 hours after posting
7. Impressions on browse feature traffic 24 hours after posting
8. Impressions on suggested traffic 24 hours after posting
9. The percentage of views coming from browse feature traffic 24 hours after posting
10. The percentage of views coming from suggested traffic 24 hours after posting
11. <u>Variables 3-10, but 72 hours after posting</u>
12. Net view total (views at 24H + views at 72H)
13. Net suggested impressions total (suggested impressions at 24H + 72H)
14. Our own metric that blended CTR with impressions to give a more accurate estimate

Results

Overall Results

As you'll see momentarily, our initial hypothesis was not fully correct. We were correct in assuming any data from the first day after a video is posted doesn't have much say on the long-term success of a video. Of course, sometimes it may feel apparent to you (i.e. you have a video performing higher than any of the last 3 months), but it won't have predictive power in your more-average videos.

We were partially correct with our theory on the role of traffic sources not being very significant. In reality, suggested traffic impressions 24 and 72 hours after a video is posted can help predict a video's success, but only in tandem with the view count. However, early view count is the most important predictor without question.

The percentage of views coming from each traffic source or their respective CTRs are *not* representative in any way (this is contrary to what we've heard many YouTube 'experts' claim).

There is no 'special' way to predict video success beyond early view totals (and for some, suggested impressions as a hint on whether the video will appeal beyond core subscribers). However, it is worth looking into specific use-cases where a bad video suddenly becomes great or vice versa down the line. These are likely not consistent patterns of something, but rather caused by a myriad of different factors.

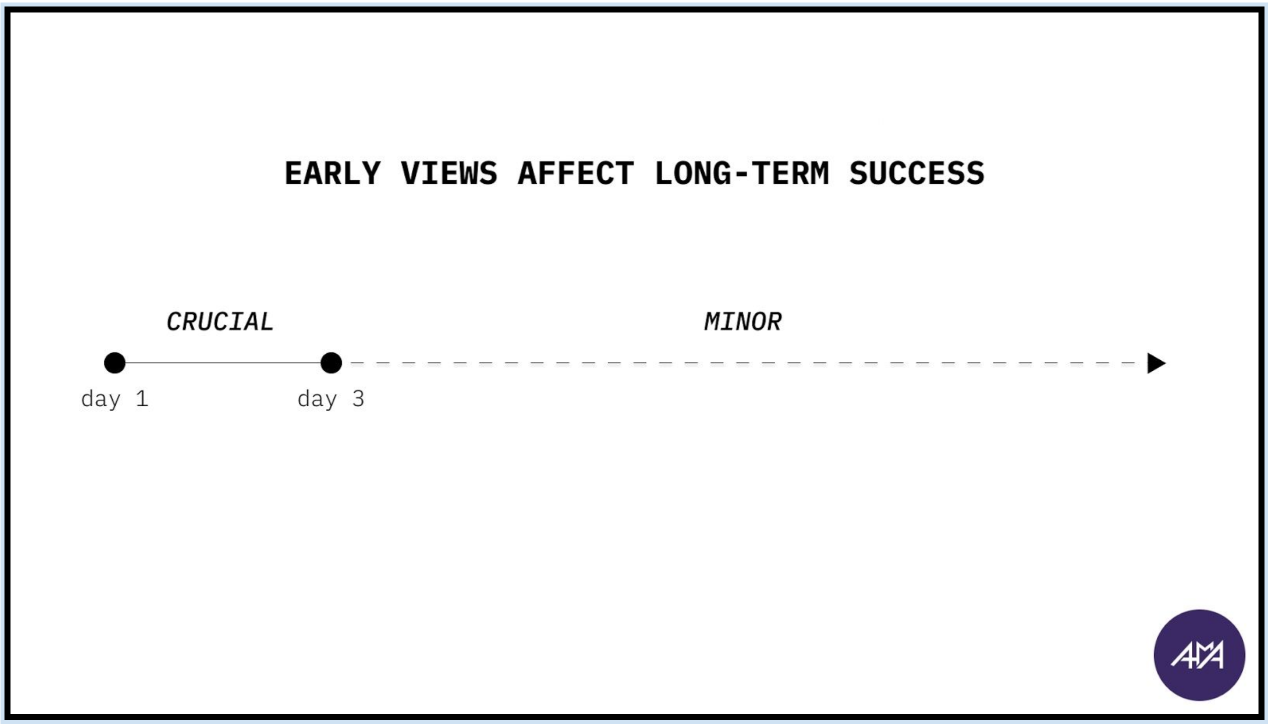
What to Expect

The next page will focus specifically on the key takeaways from the multi-linear regressions we ran with the data. Their goal is simply to outline interesting notes from the study and offer a little context on why they may exist and how it can potentially help you in the future.

Following that, in the nuanced results section, we'll look specifically at the regression data tables and scatter plots, going over which of our above metrics didn't make the cut, and further analyzing why certain patterns exist in the data.

Key Results Overview

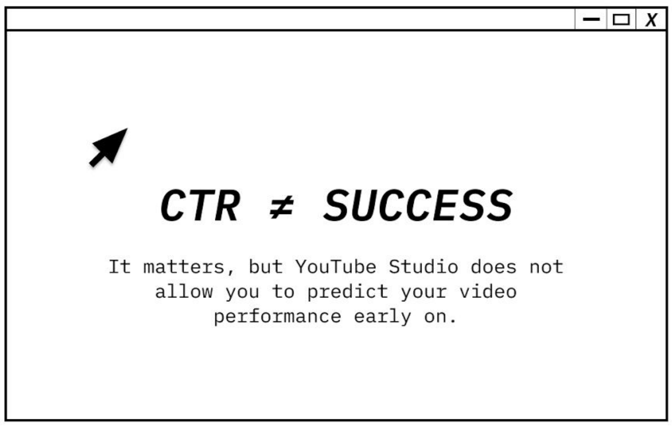
Result #1: The Amount of Views Brought in 72 Hours After Posting Disproportionately Helps Your Chances of Long-Term Growth



Not all views are created equal.

One view on your video 3-months in does not mean as much as one view on your video shortly after it's posted. We found that views coming in the first couple of days were very important in gauging the video's success rather than the view count weeks or months afterward.

Result #2: Any Information on CTR is Not That Helpful

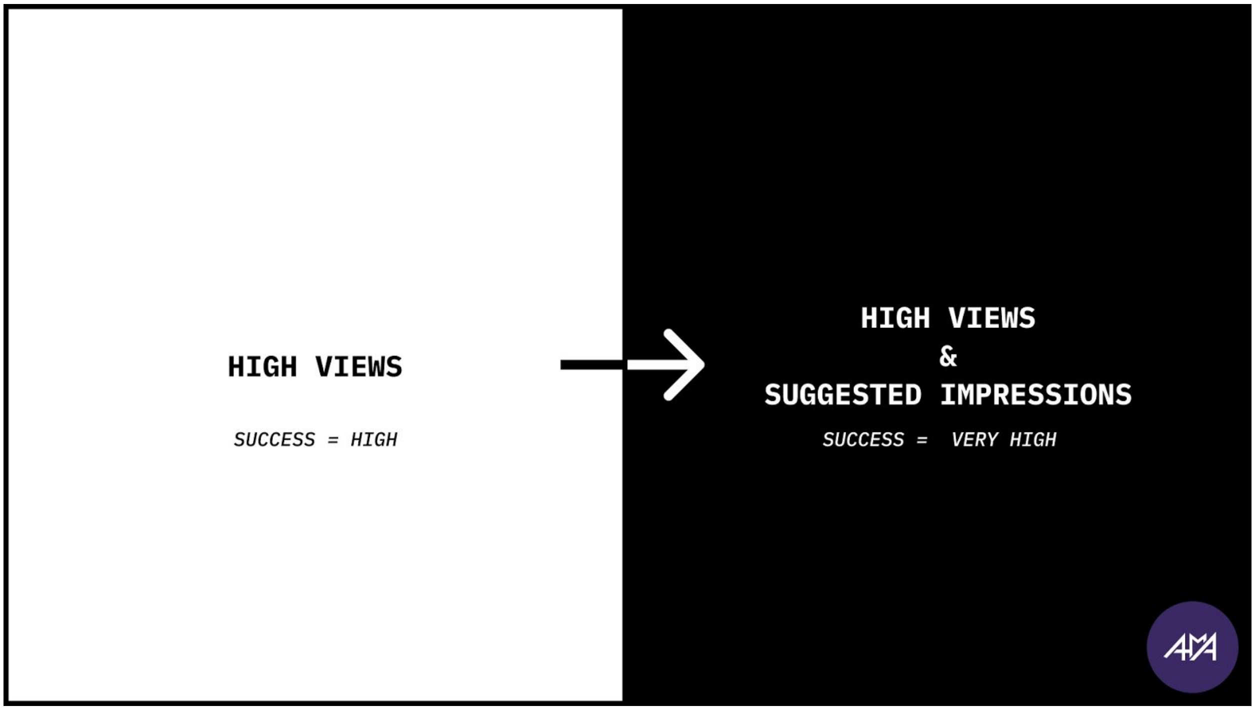


One clear result from our analyses was that clickthrough rates had very little ability to predict the success of a video. They were essentially all the variables that needed to be excluded from our model for not having enough statistical significance.

This makes sense intuitively. Think about a time when a video performing terrible had a high clickthrough rate, potentially confusing you. In reality, it's likely that this video has an artificially high clickthrough rate because the only people YouTube will even consider showing the video too are your core subscribers. You can easily picture a scenario in the reverse as well: a high-performing video having a lower clickthrough rate simply because millions of more people are being exposed to the content.

Clickthrough rate should only be looked at when also looking at the number of impressions, otherwise it is too easy to be misled by artificially low/high percentages.

Result #3: The Single-Best Predictor is Views at 72H… Depending on the Channel, When Combined With Suggested Impressions at 72H, It's Even Stronger



When in doubt, check the view-count of a video compared to your channel average, and even better, compared to the average of other videos like it you've posted (the type of content itself).

Looking at suggested impressions can potentially clue you in to whether the video will have traction with non-core fans as well. Remember: looking just after 24H isn't enough. It's much more accurate to predict a video's success after a few days.

More Nuanced Results

Data Tables and Scatter Plots

Below are the data table and scatter plot from our regression with *lifestyle creators*:

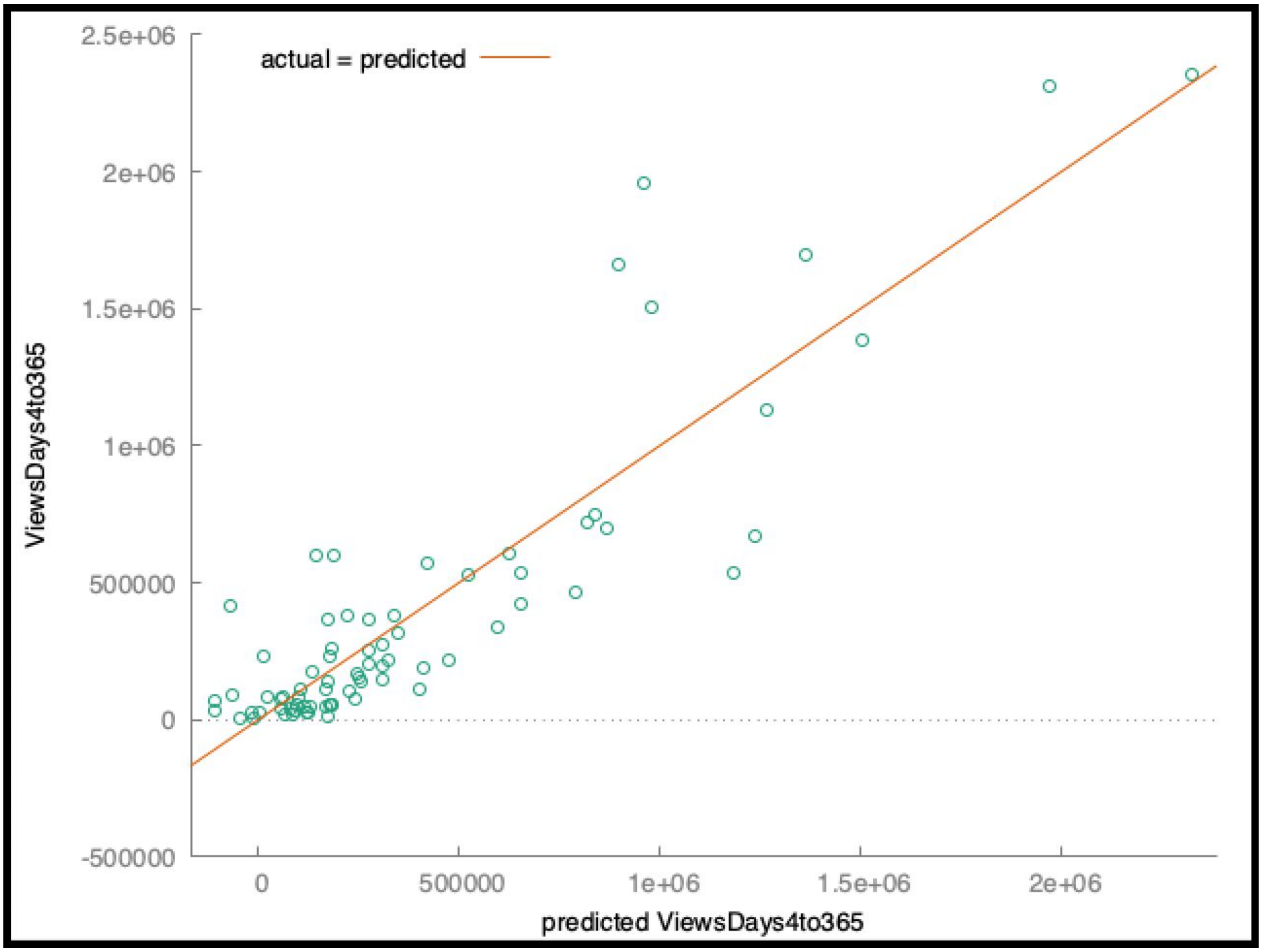
Dependent variable: ViewsDays4to365

	coefficient	std. error	t-ratio	p-value	
const	-350979	100083	-3.507	0.0008	***
Views24H	-5.37620	2.08296	-2.581	0.0120	**
SuggestedCTR24H	8.10212e+06	2.60727e+06	3.108	0.0028	***
BrowseImp24H	1.20677	0.433063	2.787	0.0069	***
SuggestedImp24H	-0.417298	0.141350	-2.952	0.0043	***
BrowseImp72H	-0.870446	0.314107	-2.771	0.0072	***
SuggestedImp72H	0.323108	0.0817021	3.955	0.0002	***
Views72H	5.40914	1.69092	3.199	0.0021	***

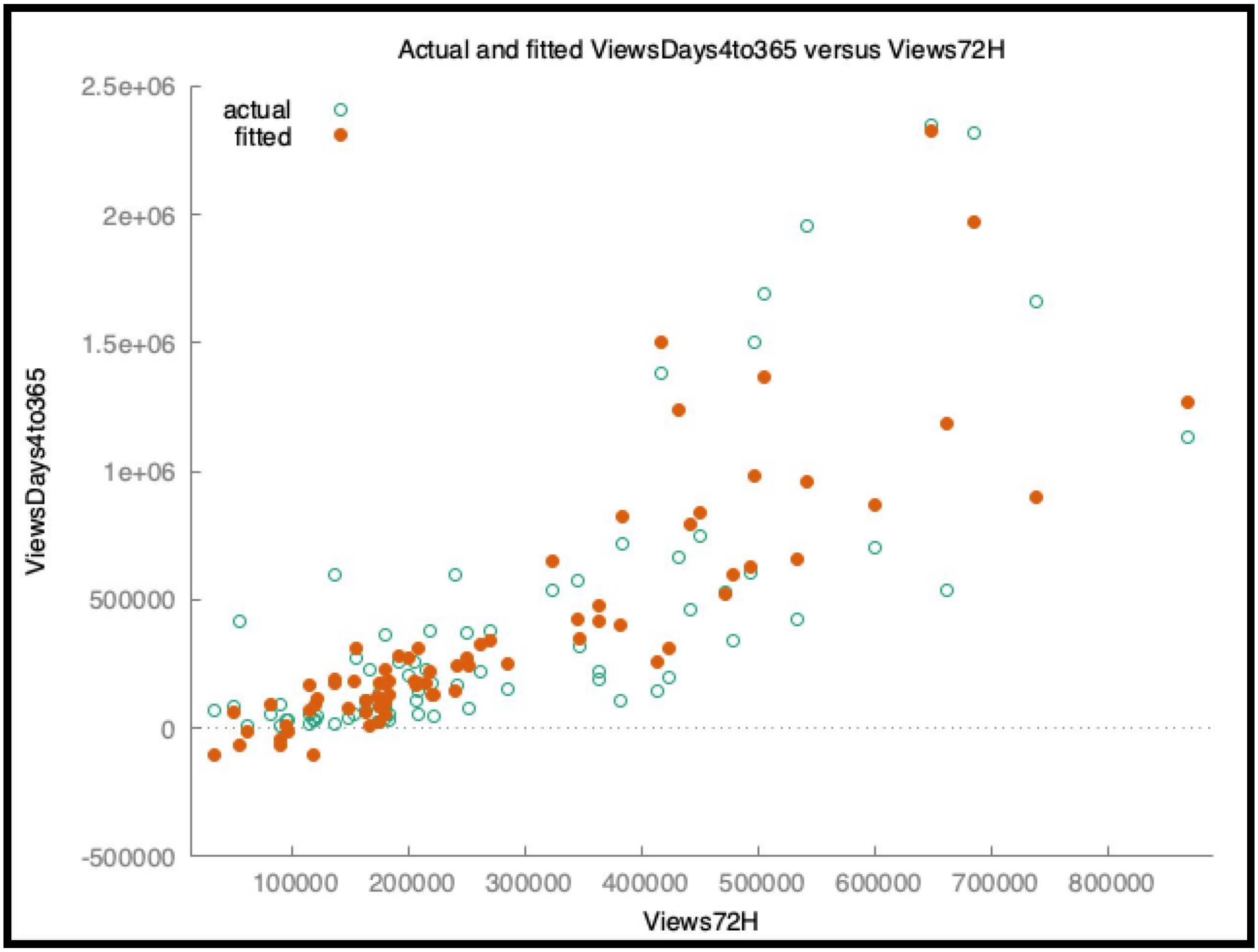
Mean dependent var	388745.0	S.D. dependent var	531402.3
Sum squared resid	4.46e+12	S.E. of regression	257946.0
R-squared	0.786669	Adjusted R-squared	0.764381
F(7, 67)	35.29516	P-value(F)	4.14e-20
Log-likelihood	-1036.728	Akaike criterion	2089.457
Schwarz criterion	2107.997	Hannan-Quinn	2096.860

If needed, a quick statistics refresher. Any variable with a p-value, seen at the top-right corner of under 0.05 means that variable is statistically significant and likely has a real correlation to the dependent variable in question (views after one year).

The below graph shows the relationship between the actual data points and the prediction generated by the model. The second graph shows specifically that relationship, but between views at 72 hours and the prediction generated by the model.



actual = predicted
ViewsDays4to365
predicted ViewsDays4to365
2.5e+06
2e+06
1.5e+06
1e+06
500000
0
-500000
0
500000
1e+06
1.5e+06
2e+06

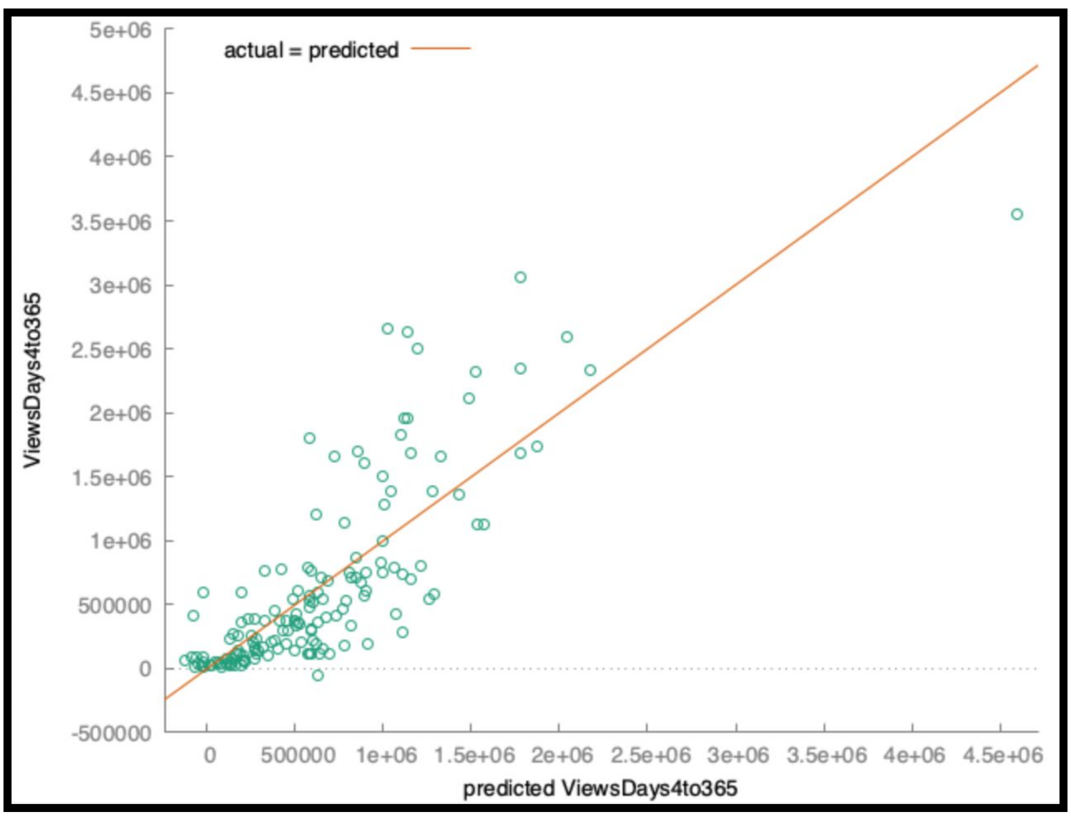

Actual and fitted ViewsDays4to365 versus Views72H
actual
fitted
2.5e+06
2e+06
1.5e+06
1e+06
500000
0
-500000
ViewsDays4to365
100000
200000
300000
400000
500000
600000
700000
800000
Views72H

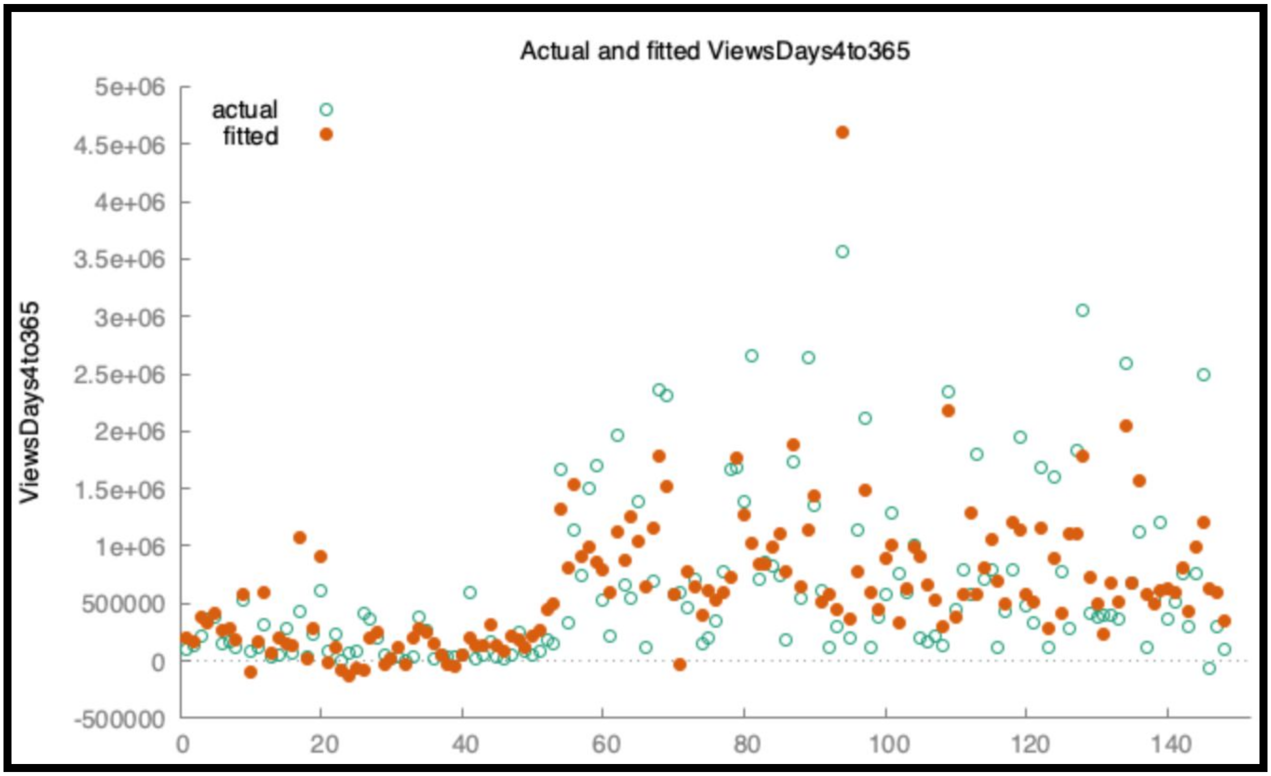
Below are the data table and scatter plot from our regression with all data:

```
Model 5: OLS, using observations 1-148
Dependent variable: ViewsDays4to365

                     coefficient     std. error    t-ratio   p-value
  --------------------------------------------------------------
  const              -190193          81086.1        -2.346    0.0204   **
  Views72H                 2.30096        0.250711    9.178    4.29e-16 ***
  SuggestedImp24H         -0.326526       0.0647721  -5.041    1.37e-06 ***
  SuggestedImp72H          0.161519       0.0294636   5.482    1.83e-07 ***

Mean dependent var    644316.1   S.D. dependent var    726581.4
Sum squared resid     2.77e+13   S.E. of regression    438519.0
R-squared             0.643176   Adjusted R-squared    0.635743
F(3, 144)             86.52027   P-value(F)            4.63e-32
Log-likelihood       -2130.667   Akaike criterion      4269.334
Schwarz criterion     4281.323   Hannan-Quinn          4274.205
```




Actual and fitted ViewsDays4to365
ViewsDays4to365
actual
fitted
5e+06
4.5e+06
4e+06
3.5e+06
3e+06
2.5e+06
2e+06
1.5e+06
1e+06
500000
0
-500000
0
20
40
60
80
100
120
140

Implications

What Now?

Our biggest takeaway from diving into the data is that you should have a holistic mindset when trying to evaluate how well a video has done after a few days. Yes, technically, views are the best predictor a few days in of the view-success of a video a year later. However, it is worth keeping other metrics, such as traffic sources, in mind.

In reality, CTR in isolation is never going to help predict video success. However, we've seen some success in blending CTR with the impression count to more accurately gauge the interest/clickthrough of a specific video. If you want a more accurate picture, we'd recommend trying that out.

Finally, this paper resolidifies a message that many YouTube "growth experts" wouldn't want you to think: there is no 'quick fix' to gaming the algorithm. In reality, continue to focus on watch time and posting consistently for the best chance at success!

Moving Forward

This paper is only the beginning of the exploration that should exist for this topic. Our data set was small (150) compared to a traditional research study on a topic like this. We encourage others to perform their own research to build on this foundation.

We will also finish up our Excel model from this study for public use. We're essentially taking the final formula from our regression, and converting it into Excel so that you can type in a couple variables and get a prediction on how well the video will perform.

With love,
The AMA Team