

## **Factors of a Successful Influencer**

Brendan Kade and Elizabeth Robles

Department of Data Analytics, McDaniel College

ANA522P: Exp Practicum Prog in Python

Dr. Xuejing Duan

May 18, 2025

**Abstract**

This study investigates the key factors that contribute to the global recognition and financial success of YouTube influencers, focusing on metrics such as subscribers, views, uploads, and earnings. Using a cleaned and filtered subset of the 2023 Global YouTube Statistics dataset from Kaggle, we analyzed 302 U.S.-based accounts through both descriptive and inferential statistics. We explored patterns across channel types and tested whether numerical variables significantly differed between them. While entertainment and music channels dominated top-ranking accounts, ANOVA tests revealed no statistically significant variation in performance metrics across channel types. A multiple linear regression model showed a moderate relationship between rank and the numerical variables, with subscriber count being the only statistically significant predictor. These findings suggest that subscriber loyalty and engagement may be more valuable indicators of influencer success than high volume of uploads or views. Companies seeking to maximize returns from influencer marketing should prioritize quality engagement over superficial metrics.

Key Words: Influencer, Subscriber, Views, Marketing

## Introduction

In our research, we aim to investigate which factors, such as views, subscribers, and income, contribute most to the global recognition and financial success of influencers. Companies have been using influencers as a popular marketing tool to get their business more recognition. In the Forbes article, “The Power Of Influencer Marketing: Your Strategic Investment For Success,” it talks about how vital it is to have influencers in marketing plans. The article states, “for every \$1 you invest in influencer marketing, you can make roughly \$6.50. Brands benefit from their cost-effectiveness in reaching specific audience segments while maximizing ROI.” (Kuzminov) This is saying that influencer marketing has excellent return on investment and that if companies want to see a rise in sales, then they should look into the right influencer for their business. However, hiring managers and marketing specialists struggle to know which media characteristics are most strongly associated with reach and profitability. In this paper, we use descriptive and inferential statistics to analyze what contributes to making a high ranking Youtube account and how companies use influencers for a great marketing tactic. We explore the value of quality over quantity of influencer’s interactions and how influencers can bring success to a variety of businesses.

## Data

In this analysis, we use data from a subset of the *Global YouTube Statistics 2023* dataset from Kaggle. The original dataset contains 995 entries and 28 variables. However, we subset the data to only seven variables. The seven key variables we use are country, rank, channel type, uploads, video views, subscribers, highest yearly earnings. Rank and channel type are categorical variables, while the rest are continuous variables. In this subset, there are no highly correlated variables. Note that we filter the observations to only US-based accounts using the country

variable. Then with only US-based accounts, we remove this variable. We also scaled down three variables (subscribers, video views, highest yearly earning) by dividing by a million, to better handle and understand the large numbers. In cleaning the data, we also detected outliers with z-scores, where we found 24 outliers. We decided to remove all outliers for the number of uploads, but only the outliers for yearly earnings and video views for accounts not in the top 100 rank because these could signify bot behavior and sponsored accounts. After cleaning the data, we have 302 observations with no missing values.

## **Methods and Analysis**

To better assess the data, we begin with descriptive statistics, including summary measures, distribution analysis, and variable correlation assessment. Specifically, we compute the mean, standard deviation, minimum, maximum, and three percentile values for the numerical variables. For categorical variables, we compare the frequencies of each channel type and visualize them using a pie chart. Noting substantial variation in channel type frequencies, we use a boxplot to investigate whether these differences influence video views. Before proceeding with inferential statistics, we evaluate multicollinearity by calculating a correlation matrix for the numerical variables and flagging pairs with a correlation coefficient above 0.9. For inferential analysis, we conduct ANOVA tests to compare numerical variables across channel types. Finally, we run a multiple linear regression to predict rank based on uploads, video views, subscribers, and highest yearly earnings.

## **Results**

From the descriptive statistics, we found that there are no highly correlated variables and all the numerical variables were right-skewed, suggesting that the quality of these factors, like uploads, is more important than the quantity. Figure 1 shows the distribution of the uploads

variables. While very low number of uploads seem to have the highest frequencies, this may be due to other factors, such as number of subscribers, having a bigger impact on influencer's success.

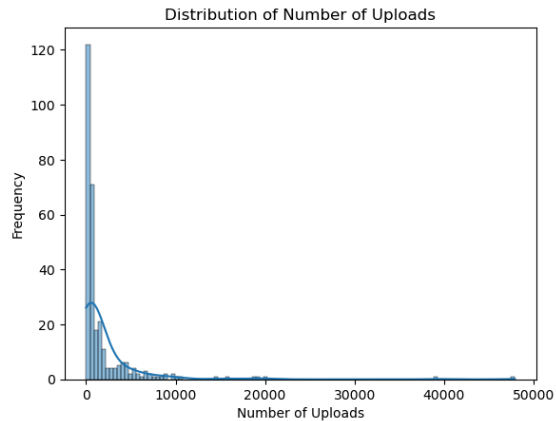


Figure 1 Histogram of Uploads

The frequencies of channel types can also provide insights to what consumers value. The pie chart in figure 2, shows that top Youtube accounts are mainly made up of entertainment (35%) and music channels (26.8%).

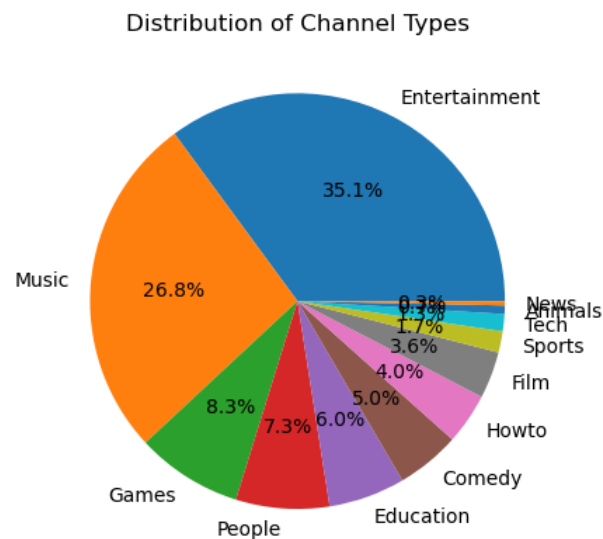


Figure 2 Pie Chart of Channel Types Distribution

Because the results of the pie chart indicate unproportional distribution of channel types, we examine the distribution of video views by channel type to determine whether channel types heavily impact other factors. Figure 3 shows the results of the box plot, which seems to show that there is no significant differences in video views between channel types.

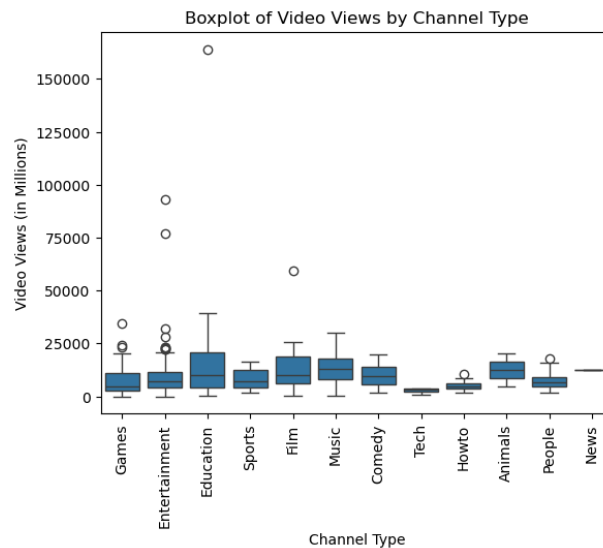


Figure 3 Box plot of Video View by Channel Type

To confirm whether there is a statistically significant difference for each of the numerical variables by channel type, we run ANOVA tests. The results of the ANOVA test show that no p-value had less than 0.05 for four variables, indicating that we do not have enough statistical evidence to reject the null that there is no difference between the channel types and the numerical variables- uploads, video views, subscribers, highest yearly earnings.

Lastly, we conduct a multiple linear regression model using rank as the dependent variables and the numerical variables as the independent variables. The results from the multiple linear regression shows a relatively weak relationship, as the adjusted R-squared is 0.368, meaning 36.8% of the variation in the dependent variable (rank) is explained by the independent variables (uploads, video views, subscribers, highest yearly earnings) in the model. However, the

results also show that subscribers are statistically significant since the p-value is lower than 0.05, meaning as subscribers decrease, rank decreases. Note that the lower the rank is, the more successful the account is. While the other variables also have negative coefficients suggesting that as factors like video views and uploads decrease, their rank decreases as well, no other variable is statistically significant. Looking specifically at the linear regression plot for number of subscribers and rank, the reason why our model was not a good fit could be because our data is not linear.

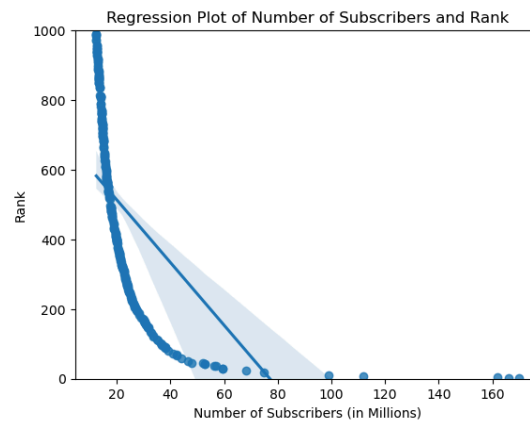


Figure 4 Regression Plot of Number of Subscribers and Rank

### Key Findings

Our research proves that the most prominent YouTube categories are Entertainment and Music, however, these categories do not directly impact the numbers of subscribers, views, uploads, and earnings. Additionally, a high number of uploads and views may not directly lead to successful influencers. For this reason, companies seeking to partner with social media influencers should prioritize creators that have many loyal subscribers and consistent engagements with viewers. An influencer with these qualities should help companies promote their brand and gain valuable customers.

## **Conclusion**

Our analysis of U.S.-based YouTube accounts using a subset of the Global YouTube Statistics 2023 dataset reveals key insights into the potential of influencer marketing. After cleaning and modifying the data, we were able to conduct a meaningful investigation into the factors driving influencer success. While entertainment and music channels are the most common among top accounts, they do not necessarily correspond with higher subscriber counts, views, or earnings. Notably, subscriber count was the only statistically significant predictor of an account's rank, highlighting the importance of audience loyalty and engagement. This suggests that companies should prioritize partnerships with influencers who maintain strong subscriber relationships rather than those with high view or upload counts alone. Such a strategy can lead to more effective brand growth and business success. Future research should consider expanding the data to include multiple social media platforms and testing additional regression models to improve predictive accuracy. As influencer marketing continues to evolve, understanding what drives true impact will be increasingly valuable for businesses.



## References

Kuzminov, M. (2024, August 13). *Influencer Marketing: Your strategic investment for success.*

Forbes.

<https://www.forbes.com/councils/forbesagencycouncil/2024/01/05/the-power-of-influencer-marketing-your-strategic-investment-for-success/>