

Social Media Influencers

Group 4 Members:

<i>Faria Tasnim</i>	<i>(1001967007)</i>
<i>Jose Acosta</i>	<i>(1001033467)</i>
<i>Masrour Naderi</i>	<i>(1001708666)</i>
<i>Sejal Bhatnagar</i>	<i>(1001951523)</i>
<i>Srusti Rajeshbhai Gajjar</i>	<i>(1001960116)</i>
<i>Suraj Suresh</i>	<i>(1002029987)</i>

Table of Contents

1. Project Overview.....3

2. Data Description3

3. Data Cleaning4

4. Research Questions6

5. Methodology6

6. Results and Discussions.....7

7. Conclusion.....19

8. Recommendation.....20

9. References.....21

Project Overview

Social media influencers have large followers and the ability to communicate directly with their followers. Given their large followings companies are eager to collaborate with influencers to market their products directly to a large and engaged audience. The report looks at 4 months of follower data for Instagram and YouTube identifying the accounts with the highest following count along with the country where the majority of each account's followers reside, category (type of content they produce), and engagement metrics such as likes. Given these attributes and metrics, the aim of the analysis through descriptive analytics is to understand the characteristic of the accounts with the largest following and observe similarities across both social media platforms. Having insight into these characteristics can help companies run successful marketing campaigns by partnering with influencers whose followers most closely resemble the demographics they aim to target.

Data Description

Dataset downloaded from Kaggle: <https://www.kaggle.com/datasets/ramjasmaurya/top-1000-social-media-channels?select=social+media+influencers+-+Instagram+in+oct.csv>

The dataset used analyzes the top 1,000 accounts across two different social media platforms, Instagram, and YouTube. The top 1,000 accounts are defined as the accounts with the highest follower count for Instagram and subscriber count for YouTube. It tracks the top accounts across 4 months, March, June, September, and October 2022. The data is available across 8 different files 4 for Instagram and 4 for YouTube. Each file represents a snapshot of the follower and subscriber count for a given month.

The Instagram files contained a total of 8 fields of which 5 are categorical and 3 are continuous, and the YouTube file contained a total of 9 fields of which 5 are categorical and 4 are continuous. See the figures below for a view of each dataset. Categorical variables are colored green and continuous variables are colored blue.

Instagram name	Name	Subscribers	Audience country	Authentic engagement	Engagement average	Category_1	Category_2
cristiano	Cristiano Ronaldo	477.9M	India	5M	6.2M	Sports with a ball	
kyliejenner	Kylie Jenner	368.1M	United States	3.5M	5.5M	Fashion	Modeling
arianagrande	Ariana Grande	329.6M	United States	2.9M	4M		

Figure 1: Instagram

Name	Youtuber	Country	Subscribers	Category_2	Avg. views	Avg. likes	Avg Comments	Category_3
T-Series	tseries	India	224.4M	Music & Dance	141.5K	4.6K	125	
Cocomelon - Nursery Rhymes	checkgate		142.5M	Education	12.5M	64.8K	0	
PewDiePie	PewDiePie	United States	111.6M	Animation	1.9M	130.5K	7.4K	

Figure 2: YouTube

Data Preprocessing

Each snapshot for both the Instagram and YouTube datasets required multiple preprocessing steps prior to analysis. The steps highlighted below were applied to both the Instagram and YouTube datasets.

1. Standardizing Column Headers

- Standardized the column headers across each snapshot file for Instagram and YouTube datasets (See figures 3 and 4 for the standardized column headers).
- Datasets were vertically stacked after column standardization.

InstagramHandle	InfluencerName	AccountCategory1	FollowerCount	AudienceCountry	AuthenticEngagement	EngagementAverage	SnapshotMonth	Ranking
instagram	Instagram	Photography	487,200,000	India	352,300	467,400	3/31/2022	1
cristiano	Cristiano Ronaldo	Sports	419,600,000	India	5,700,000	7,000,000	3/31/2022	2

Figure 3: Instagram

YouTubeName	YouTubeChannelName	ChannelCategory	SubscriberCount	AudienceCountry	AverageViews	AverageLikes	AverageComments	Month	Ranking	AuthenticEngagement
tseries	T-Series	Music & Dance	212,100,000	India	323,700	9,800	290	3/31/2022	1	10,090
checkgate	Cocomelon - Nursery Rhymes	Animation	132,100,000	United States	13,800,000	80,900	9,058	3/31/2022	2	89,958
setindia	SET India	Movies	130,400,000	India	23,600	314	21	3/31/2022	3	335

Figure 4: YouTube

- Change datatype for continuous variable from text to integer
 - Integer variables were stored as text (see figure 1, column “Engagement Average”).
 - Used Excel’s text-to-columns feature to remove the abbreviation of “M” for Millions and “K” for Thousands. This changed the datatype to an integer.
 - Multiplied the integer by either 1,000,000 or 1,000 based on the abbreviation to get the final value (see figure 3, column “Follower Count”).
- Created a column labeled “Ranking” to order the top 1,000 accounts in each month.
 - See figure 3, column “Ranking”.
- Reduced the size of each dataset from the top 1,000 accounts each month to the top 250
- Replaced NULL values for category and audience country.
 - Most accounts appeared across multiple months, and we used the category and audience country from other months to populate NULL values.

6. Combined similar categories to reduce the number of category types.
 - For example., classified sports with the ball, sports, and automotive sports as Sports.

Research Questions

Through descriptive analysis, we set forth to answer the following objectives for both Instagram and YouTube. The goal is to gain insight into the characteristics of the accounts with the largest following and observe any similarities across both social media platforms.

1. What are the top 5 countries represented in the top 250 accounts/channels for Instagram and YouTube? Are there any changes in the countries represented over time?
2. What account/channel categories are represented in the top 5 countries?
3. What is the average authentic engagement across the top 5 countries?
4. What is the engagement rate are the top 5 countries across the categories represented?
5. Of the top 10 accounts, what is the follower distribution and category makeup?
6. What is the average authentic engagement for the top 10 accounts/channels?
7. Does a higher engagement rate lead to higher accounts?

Methodology

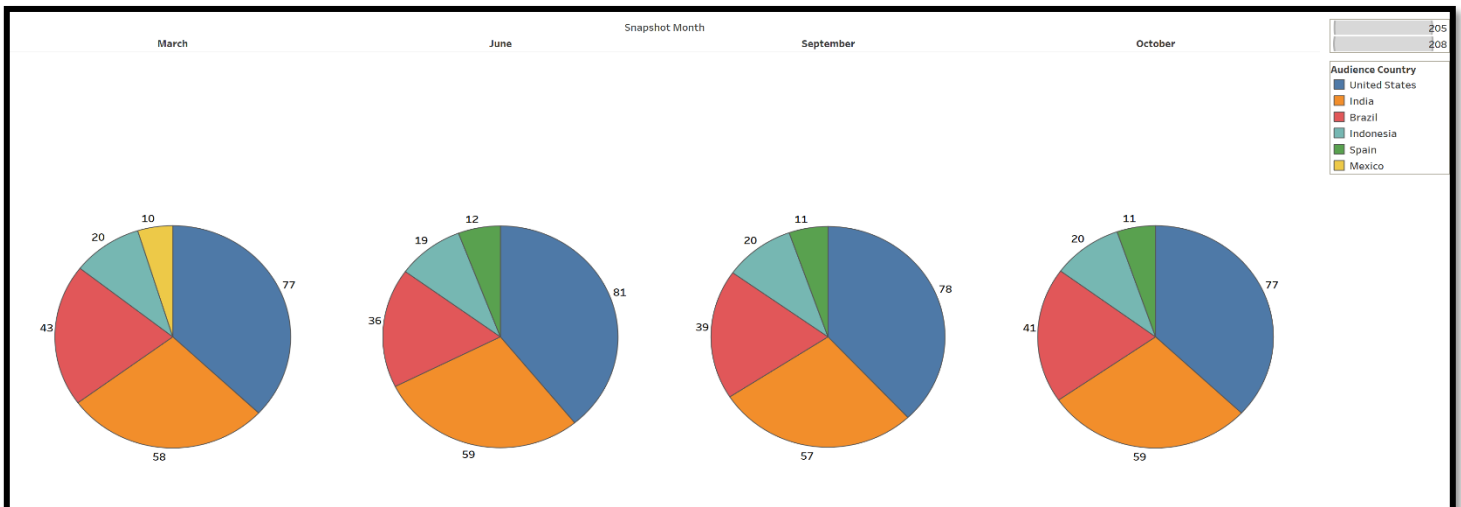
We are using descriptive analysis for our projects. A descriptive statistic is a summary statistic that quantitatively describes or summarizes features from a set of data, whereas descriptive statistics is the process of employing and analyzing those statistics.

Results and Discussion

We concentrated our attention during the analysis on two major platforms: Instagram and YouTube. We were able to compare the results and present our findings, along with interpretations, following individual analysis.

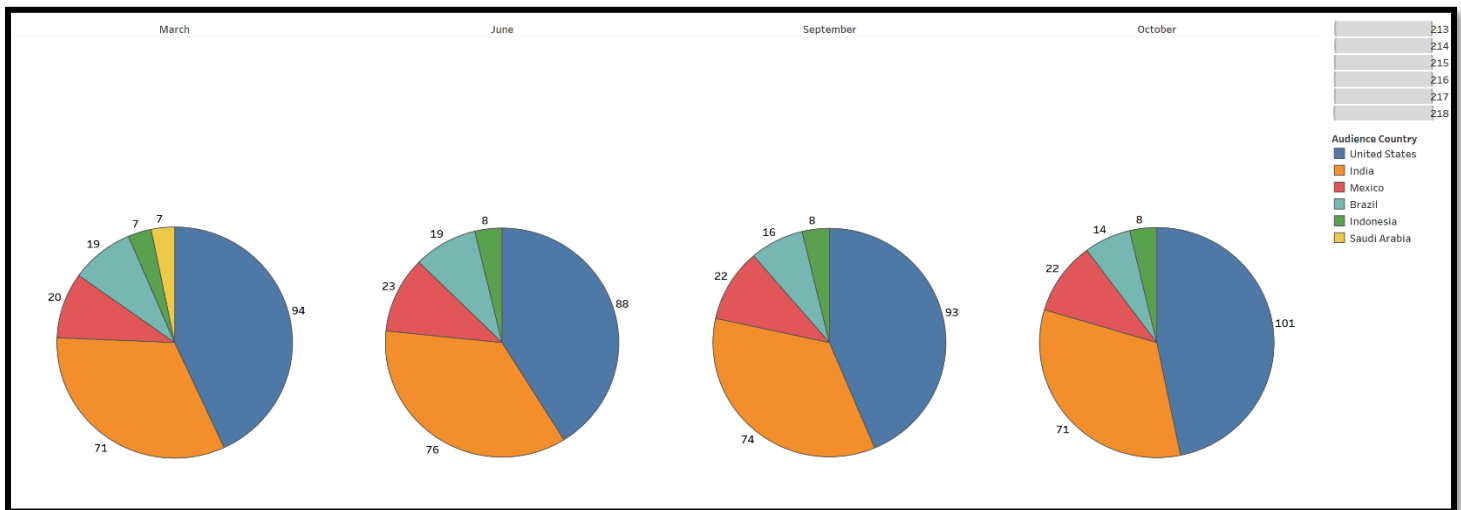
1. What are the top 5 countries represented in the top 250 accounts/channels for Instagram and YouTube? Any changes over time?

YouTube: This discusses the proportion of the top five countries by month. We can see that the top five countries on the YouTube chart stayed the same for all four months. The overall proportions also remained the same, even as the number of Americans declined in the months leading up to March. However, it increased in the months that followed. This rise can be attributed to a number of things, including a shift in social norms and more specifically.



Instagram: Similar to what we observed in the analysis of YouTube, the top five countries and their proportions remained constant across all four months for another social media platform.

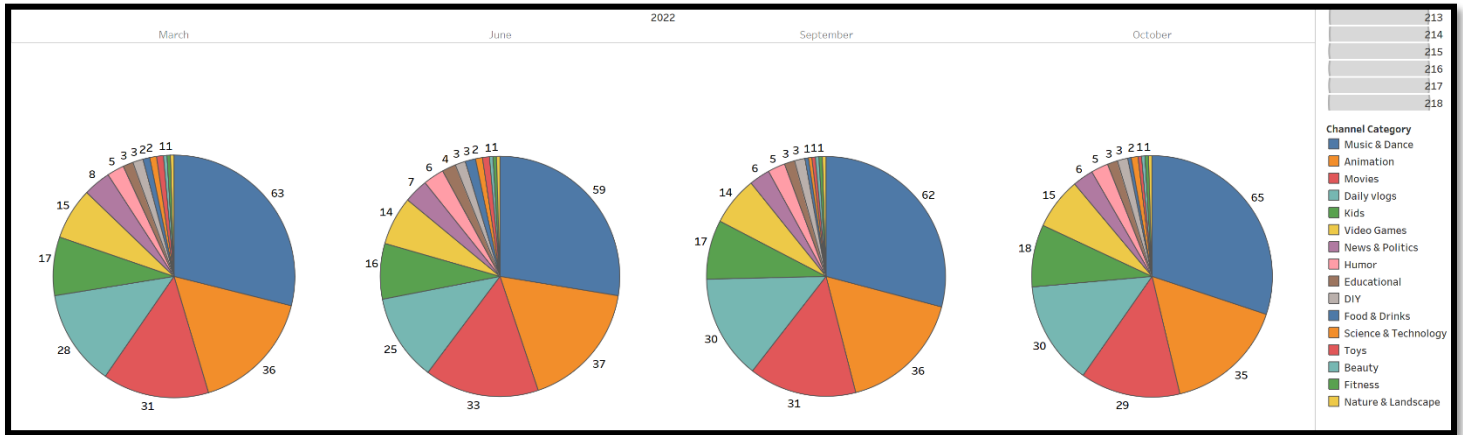
However, if we look more closely, we can see that Mexico lost ground after the first month and was replaced in the subsequent months by Spain.



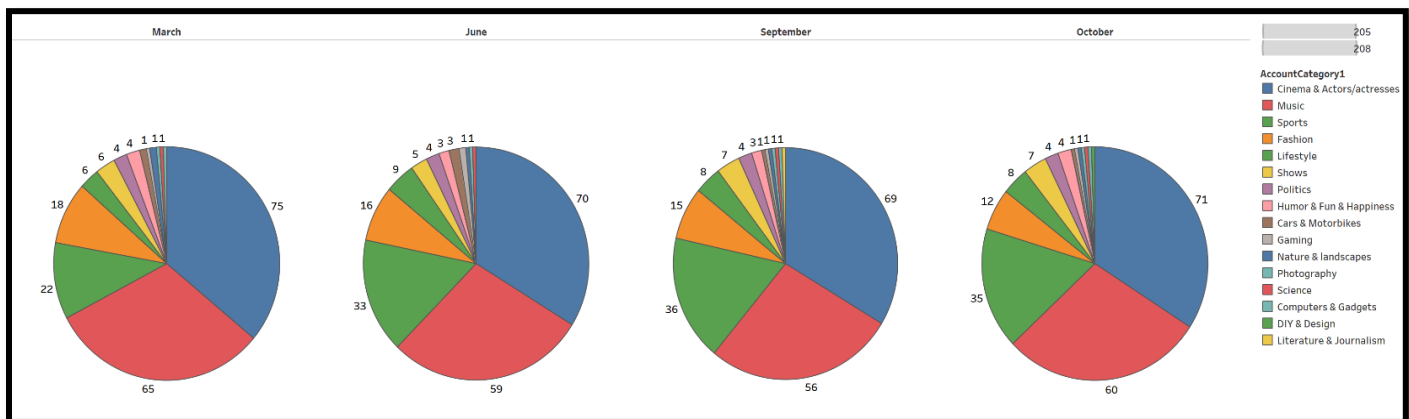
Results: There are numerous parallels that we can draw from this. First, both the platform and the top countries have significant stakes in it, and these stakes have remained constant over the months that we were looking at and analyzing. In every way, the United States, India, and Brazil have a strong reputation.

2. What account/channel categories are represented in the top 5 countries?

YouTube: For this, we attempted to examine the categories for the top five nations using the second method of analysis. Starting with YouTube, it was immediately apparent to us that the top five countries represent several account categories, the most prominent of which are Music, Dance, Animation, Movies, Daily Vlogs, Kids, and Video Games.



Instagram: The most popular categories were Film, Actors, Sports, and Fashion. There were fewer categories. That strongly suggests that many Instagram users are primarily interested in these categories.



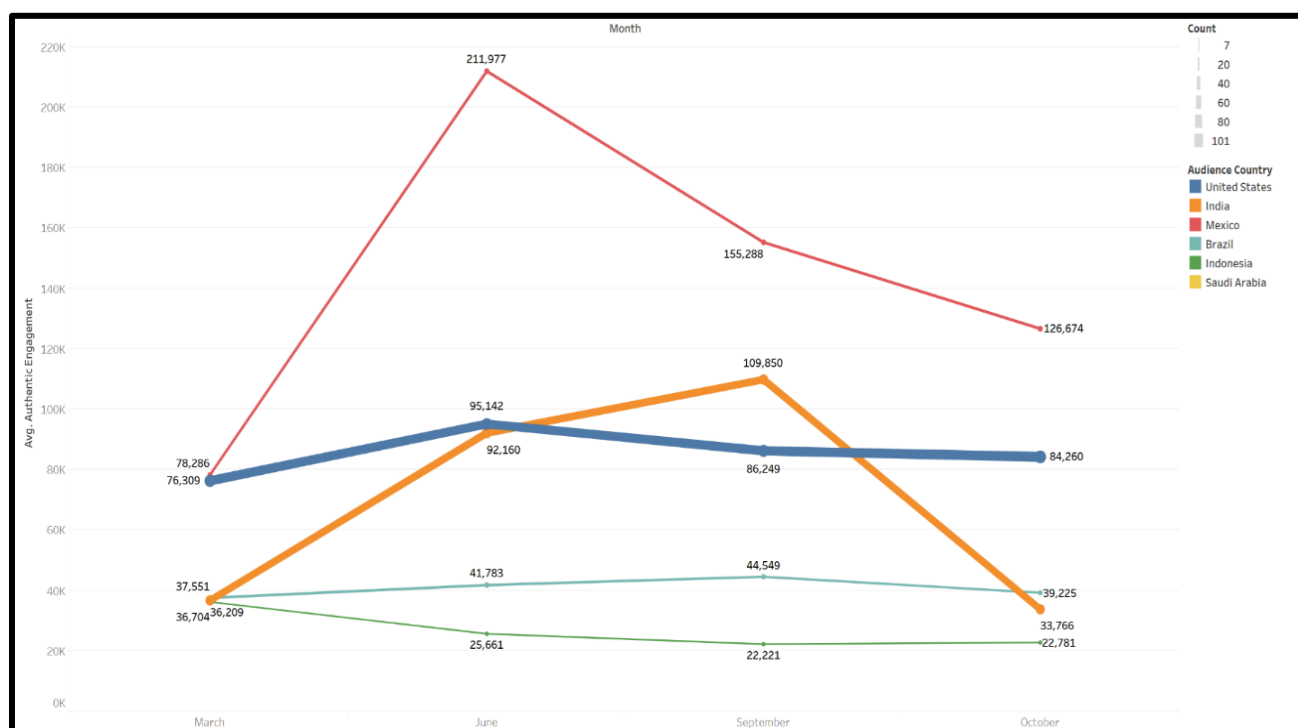
Results: Music and Cinema are among the top five categories represented on both major platforms, and there are strong contenders for both. This may be due to the recent rise in pop culture's influence and internet access as technology advances and reaches a growing number of people each day. Another reason might be that both platforms are well-known for their musical and creative filmmaking, which is why these categories typically rank highest. In addition to these top-to-strong shared categories, we also share similar weak categories across both platforms, such as Politics, DIY, video games, and news.

3. What is the average authentic engagement across the top 5 countries?

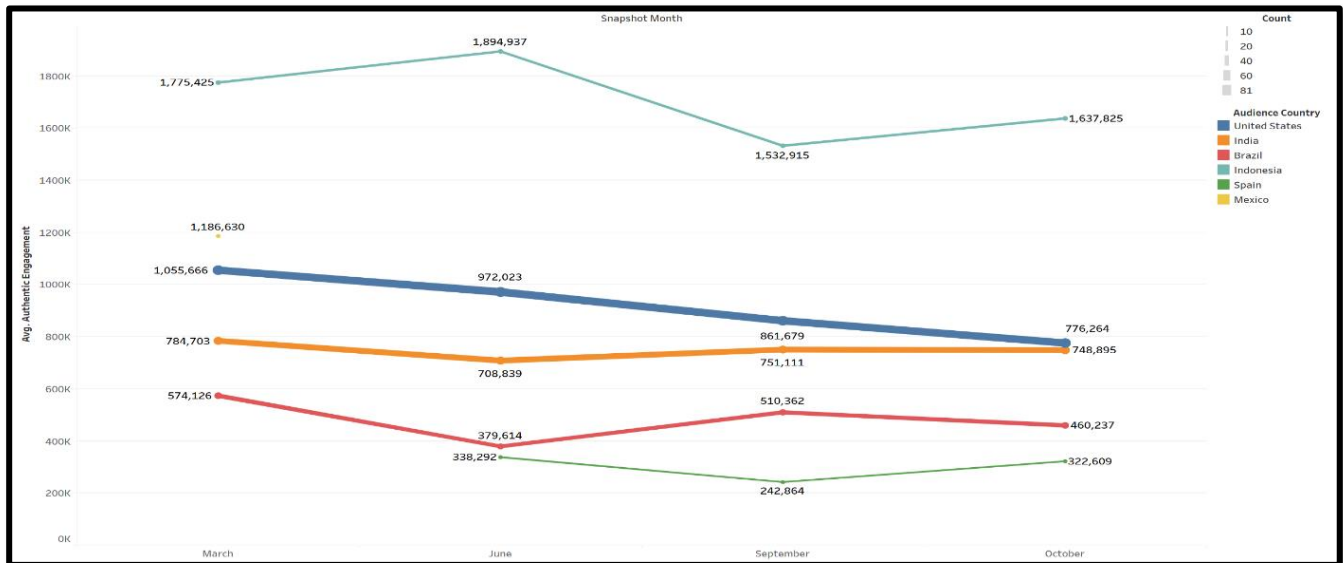
Average Authentic Engagement: The sum of the average number of likes and comments in proportion to the actual number of followers or subscribers that an account has. Just getting a subscriber or a follower on your channel or profile doesn't ensure repeat visits or retention of that user. Many users subscribe to channels as part of a campaign or promotional stunt such as giveaways. In such cases, there is very little chance that these users would revisit the channel or account for its content. Once a user has been acquired, to keep the user retained as a subscriber/follower a channel must constantly provide content that is relevant to the user and the category that the account belongs. This drives the authentic user engagement rate to get closer to the actual subscriber/follower count.

The two graphs below show the average authentic engagement amongst channels on YouTube and account on Instagram across the top 5 countries and allow us to gain a few insights from the same. The thickness of the line in the graph represents the number of accounts represented by the country, the thicker the line the higher number of countries represented.

YouTube: For this analysis overall, we have taken average authentic engagement rates per month and found that overall Mexico had the highest average authentic engagement and Indonesia had the lowest average authentic engagement. Unlike other countries, the majority of Mexico's representation was from the categories of *Animation and Daily Vlogs*. Although Mexico had the highest average authentic engagement, US and India made up most of the channels represented by the thickness of the lines.



Instagram: Similar to YouTube overall Indonesia had the highest average authentic engagement and Spain had the lowest. Indonesia's high average authentic engagement was driven by Music and Cinema accounts. Although Indonesia had the highest average authentic engagement, US and India made up most of the accounts represented by the thickness of the lines.

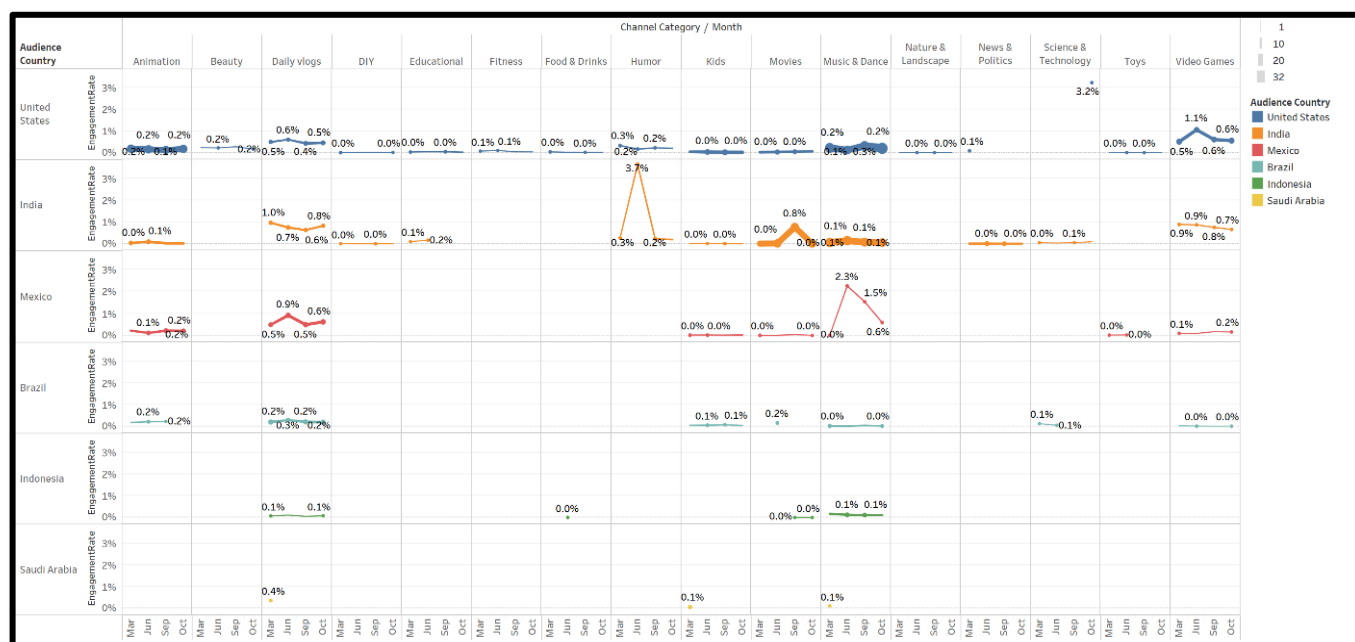


Results: Across both these platforms, Indonesia had the highest average authentic engagement for Instagram and the lowest for YouTube. This was most likely due to Indonesia's government restricting access to YouTube during this time period, which could've disenfranchised users. The maximum number of channels and accounts on YouTube and Instagram, respectively, were in the United States and India.

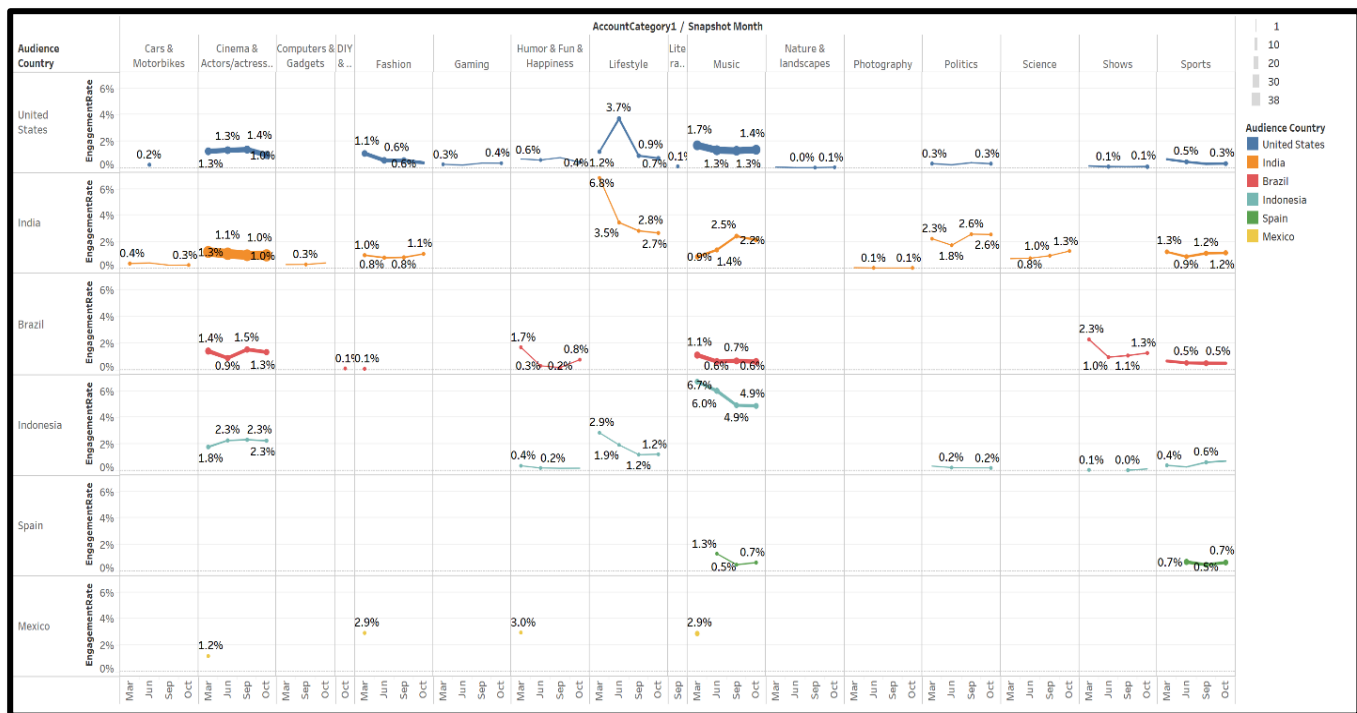
4. What is the engagement rate for the top 5 countries across the categories represented?

Engagement Rate – Authentic Engagement divided by a total follower or subscriber count. For example, if we have likes, comments, shares, and clicks on any page or a particular post or video it is authentic engagement while engagement rate is a metric often used to track how actively involved your content your audience is and how effective your brand campaigns are.

YouTube: For this, we have considered channel categories for all months with respect to audience country and found that US animation, music and dance, Video games, and Indian movies and Music and dance have the highest representation. These categories all have consistent engagement rates month over month. There is more movement in the engagement rate for less-represented categories, this is likely since one channel can have a greater impact on the overall engagement rate for less-represented categories.



Instagram: Similar to YouTube we found a Sparse matrix-like structure. US cinema and actors, US Music, and Indian cinema have the highest representation. These categories have consistent engagement rates month over month. Some of the lines are thicker than others which means the accounts represented by the country, the thicker the line the higher number of accounts represented. Also, some of the categories have spikes or sudden increases/decreases in a specific month.

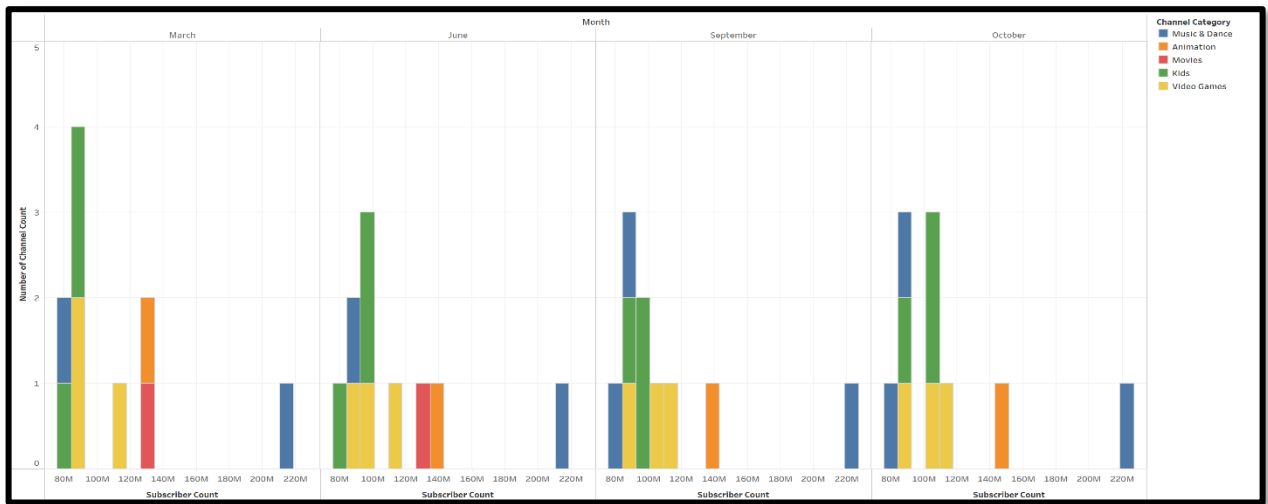


Results: From this analysis, we can say that a steady and higher engagement rate is preferred by any company to target its audiences. And here are similarities across both platforms: the prominent categories have consistent engagement rates.

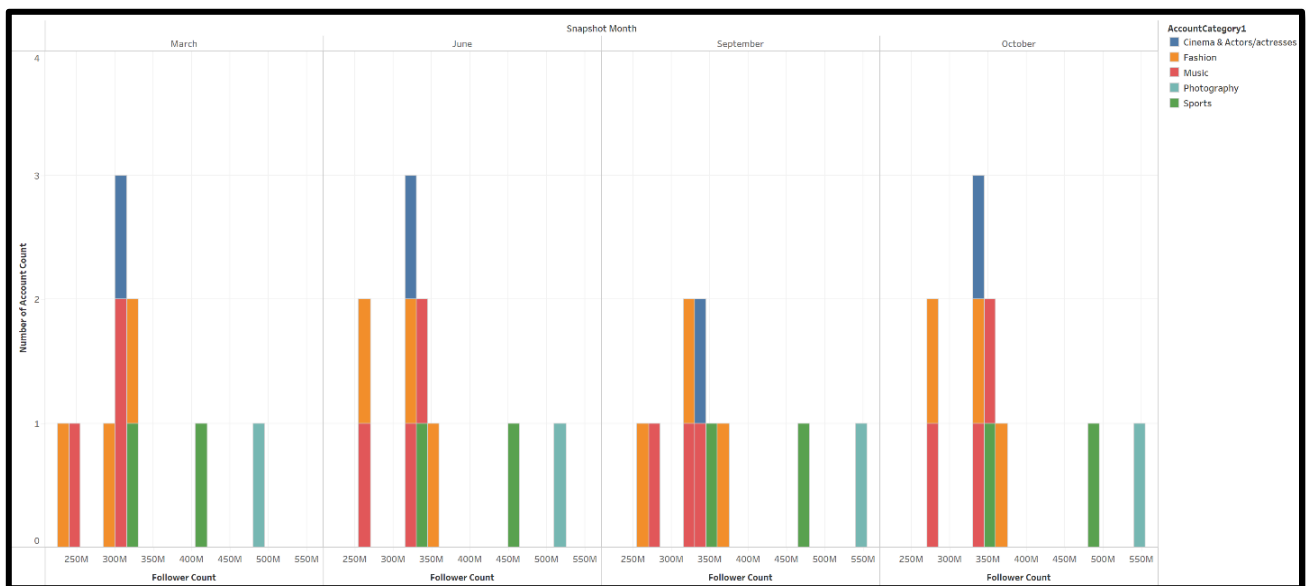
5. Of the top 10 accounts/channels, what is the follower distribution and category makeup?

YouTube (Channels): For this, we have examined the subscriber count for each month. March and October's distribution is spread out, while June's and September's distribution is tightly packed. All months have an outlier with over 200 million subscribers, from only the Music and

Dance categories. Distribution of followers means how the followers are distributed for a specific category across 4months based on the count of the followers.



Instagram (Followers): Similar to YouTube the tightly packed cluster of around 300 million subscribers across all months, each month has 2 outliers that range from 400 to 550 million followers. It is interesting to see that outliers are from only 2 categories of photography and sports, and they gradually increase over time.

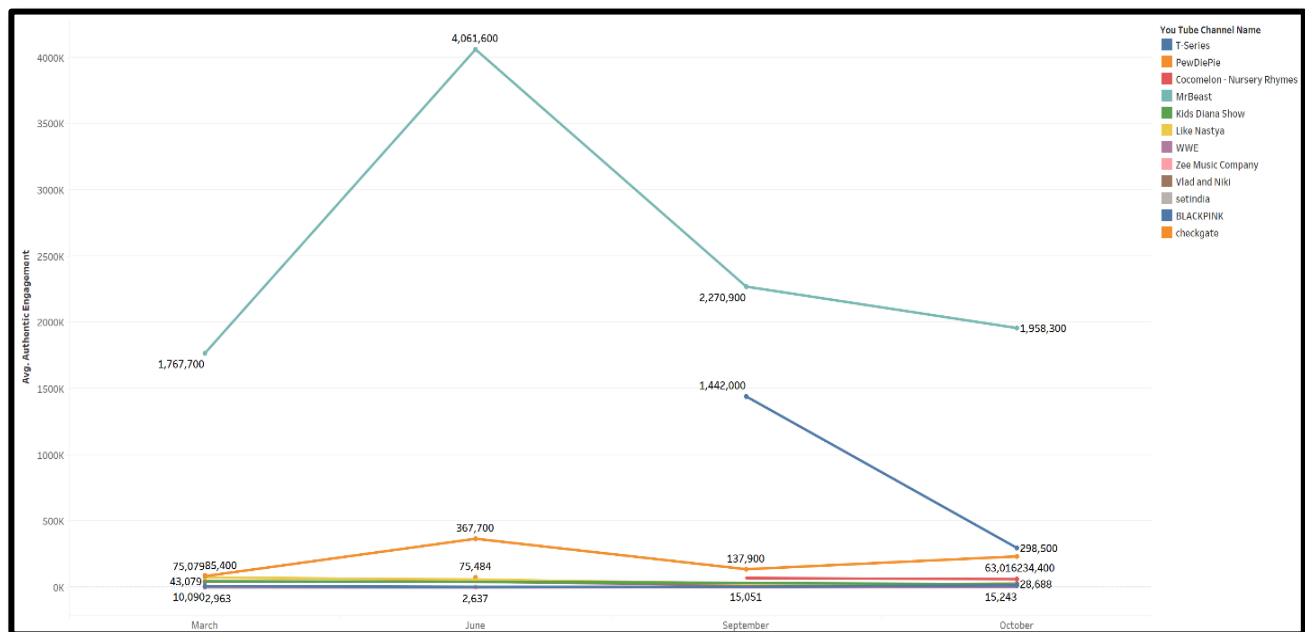


Results: Similarities across both platforms are both top 10 accounts/channels share 2 common categories in Movies and Music.

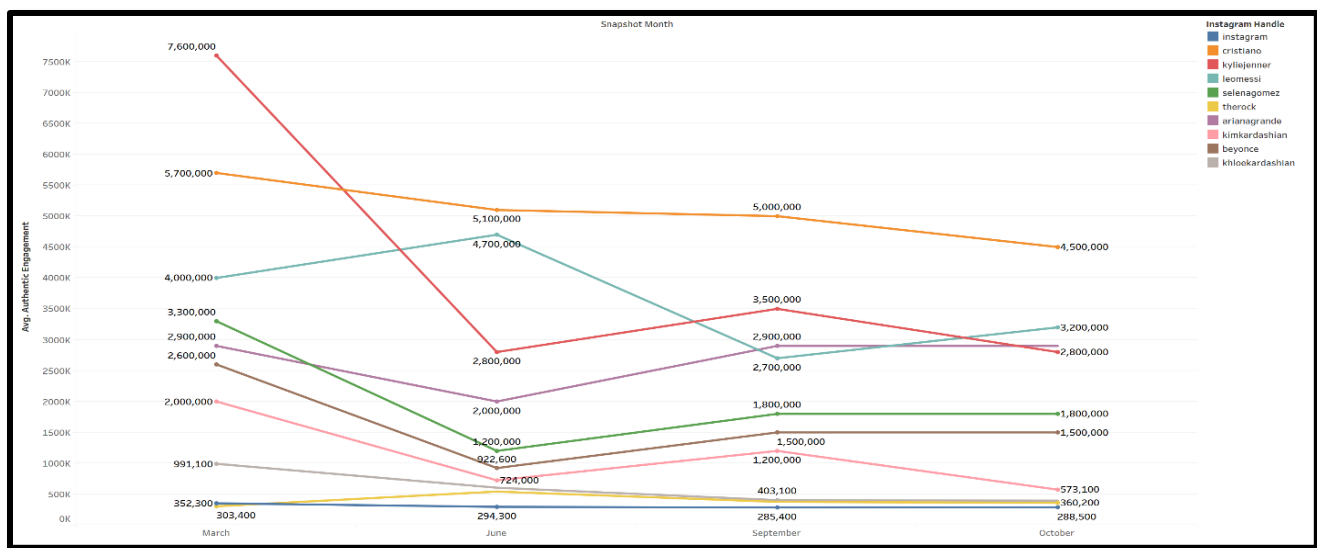
6. What is the average authentic engagement for the top 10 accounts/channels?

YouTube: Here, we have examined average authentic engagement across all months most of the top 10 channels are within the overall average authentic engagement of around 86k for the 4 months of observations. There are a few outliers in Mr. Beast, PewDiePie, and BLACKPINK.

- The content strategy of Mr. Beast is doing- learning- innovating- donating. He produces recording gaming videos on YouTube with considering millions of dollars in prizes for his followers leading to higher average authentic engagement. This is especially noted in June when he released a video Squid Game where he gave away \$500k.
- BLACKPINK's album release in September drove their average authentic engagement as the girls had been on a break for a long time.



Instagram: Similar to YouTube 7 of the top 10 Instagram accounts have an average authentic engagement for each month that is higher than the overall average authentic engagement of 1 million for the 4 months of observations. A big reduction in average authentic engagement in Kylie Jenner's account after March due to several posts from her pregnancy and the new product launch with Kendall only in March. A dip in the average authentic engagement rate in most of the accounts except for Leo Messi because of 2 reasons. First, several posts of Argentina winning the inaugural cup of champions against Italy with over 12 million likes and 150k comments. Second, his birthday' posts on Instagram gained 9 million likes and 80k comments.



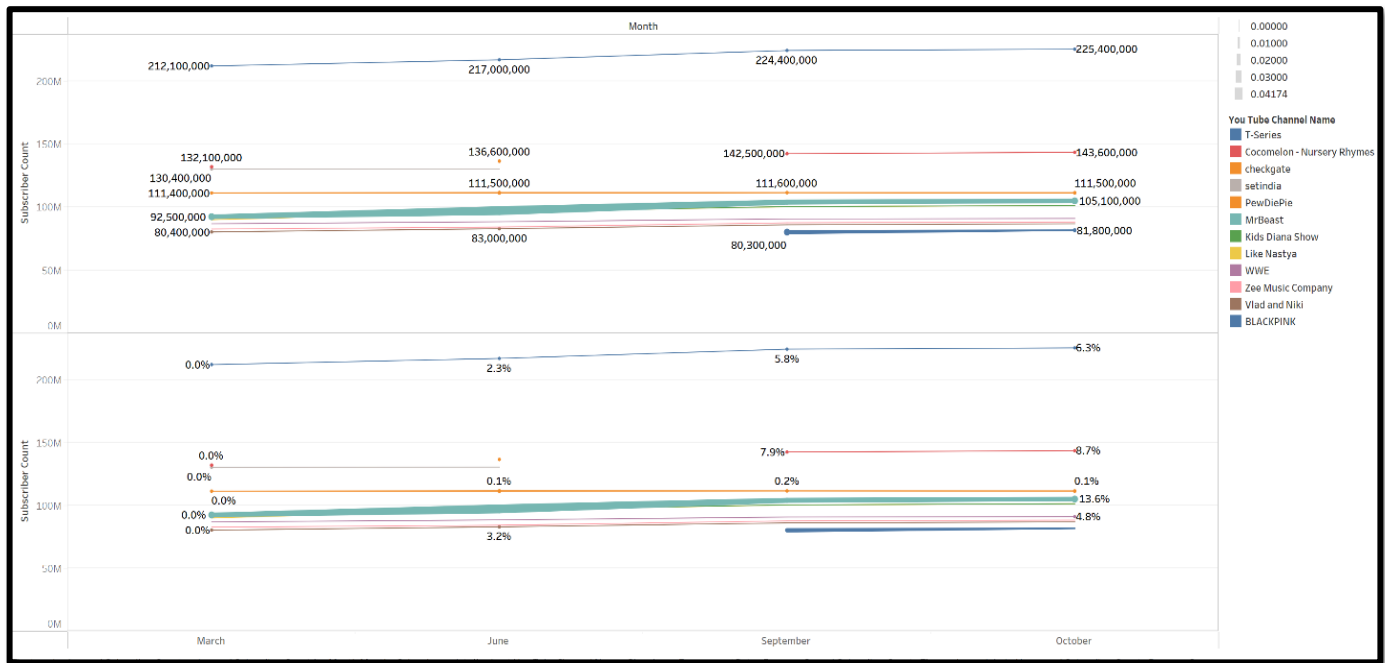
Results: Similarities across both platforms show that Average Authentic Engagement is significantly higher on Instagram than on YouTube.

7. Does a higher engagement rate lead to higher account growth rates?

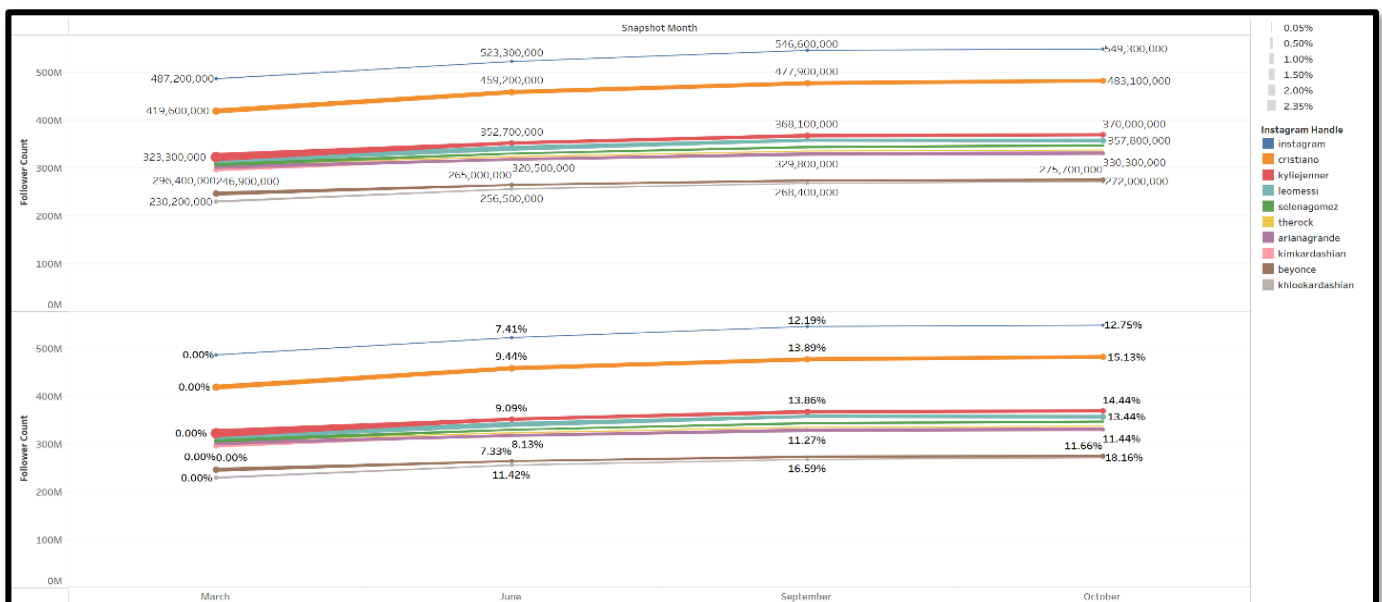
Growth Rate- Expressed as a percentage growth, measures the speed at which social media accounts are gaining or losing followers relative to their previous follower counts.

YouTube: Here from this analysis, we can say that Mr. Beast's channel has the highest growth rate in comparison to the other top 10 channels, this channel also has the highest average authentic

engagement rate. There seems to be a positive correlation between engagement rate and growth rate.



Instagram: Alike YouTube the top visualization shows the actual subscriber count, and the bottom visualization shows the growth rate of subscriber count from March to October. The Line thickness in the Follower count represented by monthly shows the engagement rate – the higher the engagement rates, the thicker the line. The cluster of accounts with around 300M followers all



have similar growth rates over the same period and these accounts belong to celebrities in cinema, music, sports, and fashion.

Results: Similarities across both platforms represent that there is a positive correlation between engagement rates and growth rates. Higher engagement rates seem to correlate to higher growth rates.

Conclusion

In conclusion, we would be summarizing our overall observations for both platforms.

Across both platforms, the top countries represented in the top 250 accounts are consistent as well as the top categories meaning the top accounts are consolidated across a few countries and categories. These top countries are the United States, India, Indonesia, Brazil, and Mexico and the common top 2 categories are movies and music for both Instagram and YouTube. Targeting audiences in these countries and categories could be beneficial for companies to reach their target customers.

Observing the authentic engagement and engagement rate for the top 10 accounts/channels we can notice that there is a positive correlation between these 2 metrics. Also, a higher engagement rate leads to a higher growth rate for a channel or an account.

Teaming up with accounts that have higher growth rates can lead to higher success rates for meeting a company's goal, as the main indicator for growth rate is engagement rate which is derived from the accounts followers' or channels subscribers' interactions.

Recommendations

Lastly, we would like to make some recommendations based on these 3 questions that we have jotted down below:

How can one use this information?

How can these findings benefit companies?

What are the drawbacks of using social media for promotion or advertising tools?

Based on the company and the type of product one is trying to sell; one can match with a channel or account that is specialized in that designated product category. Collaborate with this channel or account to promote and advertise their product to reach the largest target customer base which results in increased sales or revenue in the future. Also, word of mouth plays a significant role here.

At last, we would like to highlight our last recommendation which is not directly relevant to our analysis. But in general, everything has some pros and cons, and social media is not an exception here.

We know that there are 5 major environmental factors in any business, and they are technological factors, economic factors, social factors, political factors, and cultural factors. While we know out of the 5 factors that affect any business, political factors are the ones that create a greater impact on the level of engagement with social media and influencers. This is the only factor that also happens to be unpredictable, and therefore once implemented can have a greater impact on the engagement statistics. If a company is solely dependent on social media for advertising its product and services, then it should consider and prepare for any political conflict that may arise in the

future. Therefore, any influencer engaging with an audience on a social media platform should be prepared for any such challenges and changes in government policy.

References:

1. <https://www.kaggle.com/datasets/ramjasmaurya/top-1000-social-media-channels?select=social+media+influencers+-+Instagram+in+oct.csv>
2. <https://keyhole.co/blog/how-do-i-calculate-follower-growth-rate/#:~:text=Follower%20growth%20rate%2C%20expressed%20as,has%20seen%20since%20last%20month.>
3. <https://www.similarweb.com/blog/research/market-research/engagement-rate/>
4. <https://socialplanner.io/blog/social-media-marketing-advantages-and-disadvantages/>