YOUTUBE CONTENT STRATEGY

BY: KIRAN DAS
STUDENT ID: 24580348

SEPTEMBER —————— 2024

PROJECT OVERVIEW

The aim of this project is to leverage Snowflake, a cloud-based data warehousing platform, to create a Data Lakehouse for analyzing a dataset sourced from YouTube's top trending videos. The dataset encompasses various regions, including India, USA, Great Britain, Germany, Canada, France, Russia, Brazil, Mexico, South Korea, and Japan. It provides daily records of trending YouTube videos, including metrics like views, likes, dislikes, and comments, along with the associated category information. This initiative aims to harness the power of data to gain valuable insights, enhance data quality, and inform strategic decisions in the dynamic world of online video content.

PART 1: Data Ingestion

In this phase, the project team has:

- Downloaded and uploaded the dataset (in CSV and JSON formats) to their Microsoft Azure cloud storage account.
- Ingested the data into Snowflake as external tables.
- Transformed and loaded the data into Snowflake tables with predefined structures.
- Created a final integrated table by combining data from the trending videos and category information while generating unique IDs for each record.

PART 2: Data Cleaning

During our data cleaning process, we uncovered various issues, and we've taken steps to address some of them. Here are the outcomes:

Identifying duplicate category titles in the "table_youtube_category" table has revealed a data inconsistency where the same category title ("**Comedy**") is associated with two different category IDs. This inconsistency can significantly impact the accuracy of video categorization and analytical results. To address this issue effectively, we recommend engaging with stakeholders to reconcile and rectify the category assignments, ensuring a consistent and reliable dataset for future analyses.

During our data analysis, we identified that the category title 'Nonprofits & Activism' was unique, appearing in just one country.

Consequently, we noticed a missing category title in the final table, "**table_youtube_final.**".

The category ID of the missing title matched that of '**Nonprofits & Activism,**' as cross-referenced with the 'table_youtube_category' table. This revelation not only highlighted data uniqueness but also helped bridge the gap in our final dataset by adding the previously absent 'Nonprofits & Activism' category with the correct ID.

We discovered one video within "**table_youtube_final**" lacking a channel title- the video titled "Kala Official Teaser | Tovino Thomas | Rohith V S | Juvis Productions | Adventure Company" from India was found to have a missing channel title.

We removed **14619** records with the video ID "#NAME?" to further enhance the dataset.

We encountered duplicate records in our 'table_youtube_final' dataset. These duplicates shared the same video ID, country, and trending date, but had varying metrics such as likes, dislikes, and view counts. To ensure data accuracy, we decided to retain only the most significant information.We created a new table called 'table_youtube_duplicates' specifically for these 'bad' duplicates using a sophisticated method. We ranked the duplicates based on their view counts and retained only the records with the highest view counts, assuming that these are the most relevant.By doing this, we've not only cleaned up our dataset but also ensured that the information we're working with is the most valuable for our analyses. This meticulous process enhances the quality of our data, leading to more accurate and insightful business decisions.

PART 3: Data Analysis

This phase focuses on extracting valuable insights from the data: Found the top three most viewed videos in the "**Sports**" category for each country on '2021-10-17'.
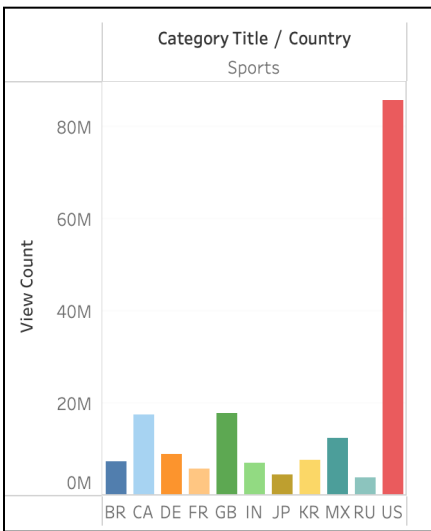

Fig 1- View Count/Category Title

The discovery of the top viewed video in the "**Sports**" category for each country on '**2021-10-17**' is noteworthy. It reveals that on this specific date, there was one standout video in the "Sports" category for each country. Notably, the United States had the highest number of views in this category, suggesting a strong interest and engagement in sports-related content within the U.S. This insight can be valuable for content creators and marketers, as it indicates a potentially large and engaged audience for sports-related videos in the U.S. on that particular day.
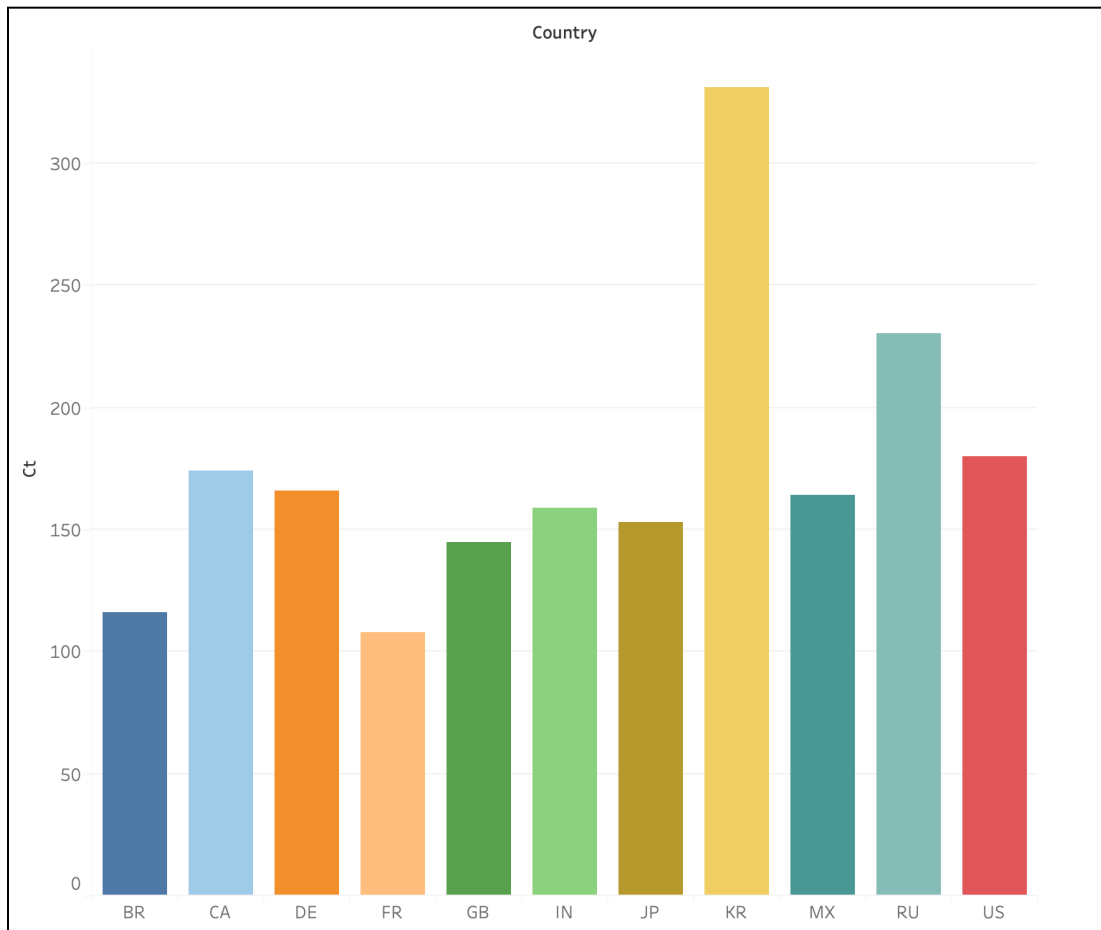
Fig 2- BTS Reference(CT)/ Country

We counted the number of distinct videos with "**BTS**" in the title for each country.The video titles in Korea, containing BTS are related to Bangtan Sonyeonda, a popular South Korean boy band. It includes various videos such as live performances, award acceptances, choreography videos, reactions, and more. For YouTube marketers, the high count of BTS-related videos in their home country suggests a strong local fanbase.

Based on the provided data, we've determined the most viewed video for each country on a

monthly basis and calculated the likes ratio, which indicates the percentage of videos in the top category compared to the total number of videos for each country. Here's one of our findings:

ChannelTitle	BR	CA	DE	FR	GB	IN	JP	KR	MX	RU	US
5-Minute Crafts FAMILY										Abc	
(G)I-DLE (여자)아이들 (Offic..				Abc			Abc	Abc			
AdeleVEVO	Abc	Abc	Abc	Abc	Abc		Abc	Abc	Abc		Abc
amazon		Abc	Abc	Abc							Abc
Bad Bunny	Abc								Abc		
BANGTANTV	Abc			Abc				Abc	Abc		
Big Hit Labels	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
BlackEyedPeasVEVO	Abc	Abc	Abc	Abc	Abc				Abc		Abc
BLACKPINK	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
Daft Punk									Abc		
Dan Rhodes					Abc						Abc
dednahype										Abc	
DisneyMusicVEVO				Abc							
DJSnakeVEVO									Abc		
DM - Desi Melodies						Abc					
Dushyant kukreja						Abc					
FAMILY BOOMS		Abc									
FFUNTV	Abc										Abc
Fortnite Fun TV											Abc
Google									Abc		
Heroindori						Abc					
Hombale Films						Abc					
How Ridiculous			Abc							Abc	
HYBE LABELS	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
Itzshauni			Abc		Abc						
Jeep					Abc						
Julius Dein		Abc	Abc		Abc				Abc		
JYP Entertainment	Abc			Abc			Abc	Abc			
Kep1er							Abc				
LankyBox										Abc	
League of Legends	Abc							Abc	Abc		
LilNasXVEVO	Abc			Abc					Abc		
MrBeast		Abc	Abc		Abc	Abc					Abc
Paytm						Abc					
Peep Peep						Abc					
RicardoMontanerVEVO		Abc							Abc		Abc
Romina Gafur										Abc	
Shanger Danger					Abc						
SMTOWN	Abc			Abc			Abc	Abc			
SportsNation											Abc
T-Series			Abc			Abc					
Tekashi 6ix9ine										Abc	
TheWeekndVEVO	Abc										
Tips Official						Abc					
Tsuriki Show		Abc	Abc		Abc						
Warner Bros. Pictures					Abc						
Wood Mood		Abc									
Woody & Kleiny Extra		Abc	Abc			Abc					
YouTube										Abc	
Алексей Навальный										Abc	

Fig 3- ChannelTitle/ Country

Several channels, such as Black Pink and Big Hit Labels, have consistently maintained high

rankings in almost all countries.

Title	BR	CA	DE	FR	GB	IN	JP	KR	MX	RU	US
6IX9INE- TUTU (Official M..										Abc	
8 Year Old Loses His Mind!..						Abc					
(여자)아이들((G)I-DLE) - '화(..				Abc			Abc	Abc			
😱OMG😱 BEST TEACHER										Abc	
$456,000 Squid Game In R..		Abc	Abc		Abc	Abc					Abc
Adele - Easy On Me (Offici..	Abc	Abc	Abc	Abc	Abc		Abc	Abc	Abc		Abc
Amazing Liquid Experime..		Abc									
Amazon's Big Game Com..		Abc	Abc	Abc							Abc
BAD BUNNY x ROSALÍA - L..	Abc								Abc		
Balloon Pop Racing Is INT..			Abc							Abc	
Black Eyed Peas, Shakira -..	Abc	Abc	Abc	Abc	Abc				Abc		Abc
BLACKPINK - 'Ice Cream (..	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc		Abc
BLACKPINK – 'Lovesick Gir..	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
BTS (방탄소년단) 'Butter' O..	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
BTS (방탄소년단) 'Dynamite..	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
BTS (방탄소년단) 'Film out' ..							Abc	Abc			
BTS (방탄소년단) 'Life Goes ..	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
BTS (방탄소년단) 'Permissio..				Abc			Abc	Abc	Abc		
Daft Punk - Epilogue										Abc	
Dance your 💜 out to #Pe..									Abc		
Diver Cracks Egg at 45 ft ..					Abc						
DIY hands cast #Shorts		Abc									
DJ Snake, Ozuna, Megan T..									Abc		
Don't be Oversmart 😎 ~ ..						Abc					
Exercise Ball Racing Is AW..										Abc	
Filhaal2 Mohabbat	Aksh..						Abc					
FUNNY DIY FOOT SOAP ◆..										Abc	
Ghareeb Baap aur Khilone..						Abc					
Godzilla vs. Kong – Officia..					Abc						
Google — Year In Search 2..									Abc		
Guess the sound challeng..		Abc	Abc		Abc						
He went against a pro and..											Abc
Husnn Hai Suhaana New - ..						Abc					
I broke my finger! 😨 (Beh..											Abc
Img it popped!! #Shorts										Abc	
Insaniyat shayad ab bach..						Abc					
Jeep®	The Middle					Abc						
JETSKI WAX PRANK!! 🤣 -..		Abc	Abc		Abc					Abc	
Join the BTS #Permission..										Abc	
Kep1er 케플러	'WA DA DA'..							Abc				
KGF Chapter2 TEASER |Ya..						Abc					
Let's just have fun with y..	Abc			Abc					Abc		
Lil Nas X - MONTERO (Call ..	Abc			Abc					Abc		
LISA - 'LALISA' M/V	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc

Fig 4- VideoTitle/ Country

Similarly most of the BTS songs are a hit in several countries.

A question was asked about the most popular content category (category_title) for each

country, based on the number of distinct videos. Here are some insights based on the obtained results:

	COUNTRY	CATEGORY_TITLE ⋯	TOTAL_CATEGORY_VIDEO	TOTAL_COUNTRY_VIDEO	PERCENTAGE
1	BR	Entertainment	4,293	16,371	26.22
2	CA	Entertainment	4,313	20,807	20.73
3	DE	Entertainment	6,679	25,299	26.40
4	FR	Entertainment	5,297	22,096	23.97
5	GB	Entertainment	4,511	20,472	22.04
6	IN	Entertainment	12,839	29,431	43.62
7	JP	Entertainment	4,945	14,816	33.38
8	KR	Entertainment	4,625	13,457	34.37
9	MX	Entertainment	3,628	15,347	23.64
10	RU	People & Blogs	10,400	63,877	16.28
11	US	Entertainment	3,812	19,130	19.93

Fig 5-Distinct Category Videos/ Country

Entertainment Dominates in Many Countries: In several countries **(BR, CA, DE, FR, GB, IN, JP, KR, and MX**), the "Entertainment" category has the most distinct videos. This suggests that entertainment content is widely popular across these diverse regions.Diversity in Content Preferences: It's interesting to note that **Russia (RU)** stands out with the "**People & Blogs**" category having the most distinct videos. This indicates a different content preference compared to other countries, where entertainment is more dominant.Varying Percentage Values: The percentage values vary significantly from country to country. For instance, in **India (IN)**, the "**Entertainment**" category accounts for a substantial **43.62%** of the total distinct videos, indicating a strong preference for this type of content. On the other hand, in Russia (RU), the "**People & Blogs**" category comprises only **16.28%** of the total distinct videos.

The channel title "**Colors TV**" has produced the most distinct videos, totaling **805** videos. This information indicates that Colors TV is a prolific content creator on YouTube.

This is what it tells us about the Channel:

- **Content Volume:** Colors TV's significant production of 805 distinct videos suggests a consistent and active presence on the platform. This can be a strategic move to engage with their audience regularly and maintain viewer interest.
- **Diverse Content:** To have such a high number of distinct videos, Colors TV likely produces a wide range of content, catering to various viewer preferences. This diversity can help them appeal to a broader audience.
- **Viewer Engagement:** Consistently creating content can contribute to increased viewer

engagement and subscriber loyalty. Frequent updates keep the channel relevant and maintain viewer interest over time.

- **Potential for Monetization:** With a substantial video library, Colors TV may have opportunities for monetization through advertising revenue, sponsorships, or collaborations.
- **Competition:** In a competitive digital entertainment landscape, the ability to produce a large number of distinct videos reflects a commitment to staying at the forefront of the industry and staying ahead of competitors.

HYPOTHESIS: Starting a new YouTube Channel

Engagement vs. Quantity: It's crucial to understand that having the highest number of trending videos in a category for a given month doesn't necessarily translate to having the highest levels of engagement. Quantity doesn't always equate to quality in terms of user interactions.

Ranking vs. Views: A category might consistently rank at the top across several months, but another category that ranks lower may actually accumulate more views. This highlights the importance of looking beyond rankings and considering actual viewership.

Likes vs. Dislikes Trade-off: Even if a category dominates in terms of the number of trending videos in a specific month, it may not have the highest positive engagements, which includes metrics like likes, shares, and comments. Therefore, relying solely on the count of trending videos per category can be misleading.

Engagements Aggregate Approach: To gain a comprehensive understanding, we can aggregate engagement metrics, such as (Views + Likes - Dislikes), for all categories each year. This approach considers the overall appeal of videos within each category over time, accounting for seasonality. It allows us to assess the category's performance more holistically, recognizing that a category may have fluctuations in its performance in different seasons.

Category Performance Year Wise
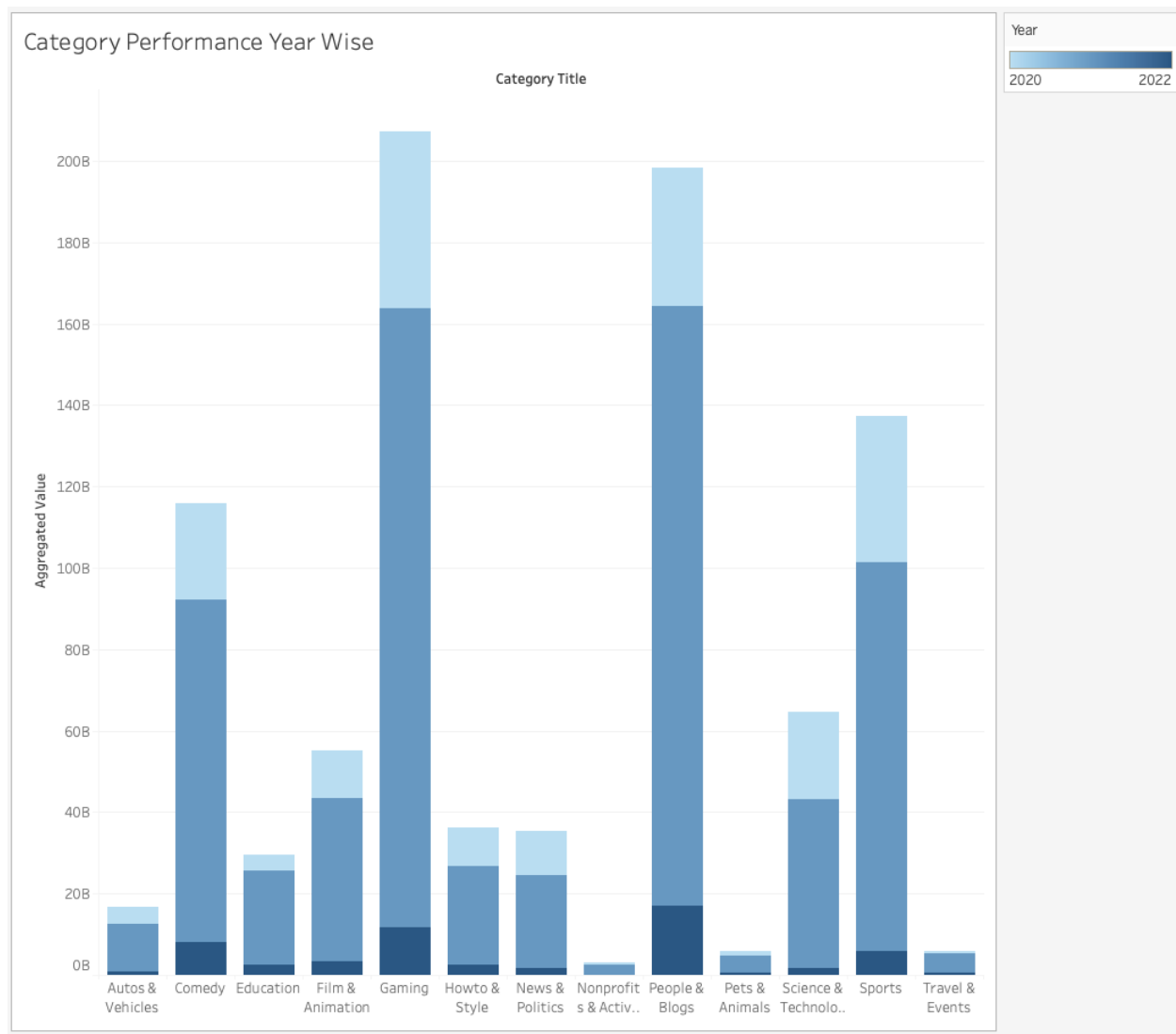
Here are some year-over-year (YoY) insights on the top 4 categories based on the data we gathered:

The aggregated metric (Views + Likes - Dislikes) is a good indicator of overall engagement with YouTube videos. It provides insight into the net positive interaction with the content, accounting for both positive (likes) and negative (dislikes) feedback, along with the number of views. Here is our analysis: :

People & Blogs Category: In the "People & Blogs" category, the aggregated value increased from 30.95 billion in 2020 to a whopping 142.53 billion in 2021. This indicates that content creators in this category received a significantly positive response from viewers, with a substantial increase in likes compared to dislikes.

Gaming Category: The "Gaming" category had a remarkable increase in aggregated value, going from 38.73 billion in 2020 to a staggering 152.05 billion in 2021. This indicates that gaming videos attracted a massive positive response from viewers, with an overwhelming increase in likes compared to dislikes. It's a strong indicator of the popularity and engagement with gaming content.

Comedy Category: In the "Comedy" category, the aggregated value increased significantly from 21.73 billion in 2020 to 84.12 billion in 2021. This suggests that not only did the number of views increase, but the net positive interaction (likes minus dislikes) also grew substantially. Viewers engaged more positively with comedy content in 2021, resulting in a higher aggregated value.

Education Category: The "Education" category also saw substantial growth in aggregated value, rising from 3.83 billion in 2020 to 24.38 billion in 2021. This suggests that educational videos not only garnered more views but also received a significantly higher number of likes compared to dislikes. Viewers found educational content highly engaging.

CONCLUSION

To determine the best category to launch a new YouTube channel that could potentially appear in the top trends of YouTube, we should consider the two categories with the highest aggregated values (Views + Likes - Dislikes). These categories tend to attract high engagement and positive interactions from viewers.

In the provided data, the two categories with the highest aggregated values are:

Gaming (2021): Aggregated Value - 152.05 billion

People & Blogs Category: (2021): Aggregated Value - 142.12 billion

Launching a YouTube channel in either the "Gaming" or "People & Blogs Category" category could be a promising strategy, as these categories have demonstrated significant viewer engagement and positive interactions.

Here are some considerations:

Gaming Category:

Gaming content has a global audience and can attract a dedicated and passionate fan base. High engagement and interaction are common in this category. But on the other hand, the

competition in the gaming category can be fierce, with many established channels.

People & Blogs Category:

Creating diverse content is key in the "People & Blogs" category, which offers a wide range of content possibilities, from vlogs to personal stories. This versatility attracts a diverse audience and fosters a strong personal connection between content creators and viewers. Focusing on the "Gaming" and "People & Blogs" categories can be a successful strategy in many countries, given their general popularity based on aggregate data. However, it's crucial to remain adaptable, as YouTube trends can shift due to evolving user preferences, emerging creators, and changing algorithms.

While these categories have shown popularity across countries, it's not a guarantee of success in every region. Local factors such as language, culture, and specific regional interests can influence viewership. Therefore, while these categories provide a strong starting point, content creators should always consider localizing their content and staying attuned to audience feedback to adapt their strategies effectively in different countries.

REFERENCES

1. R. Zhou, S. Khemmarat, and L. Gao, "The impact of YouTube on recommendation system on video", IMC New York, pp. 404-410, 2010
2. K. Filippova, and K. B. Hall,"Improved video categorization from text metadata and user comments", SIGIR New York,pp. 835-842, 2011.