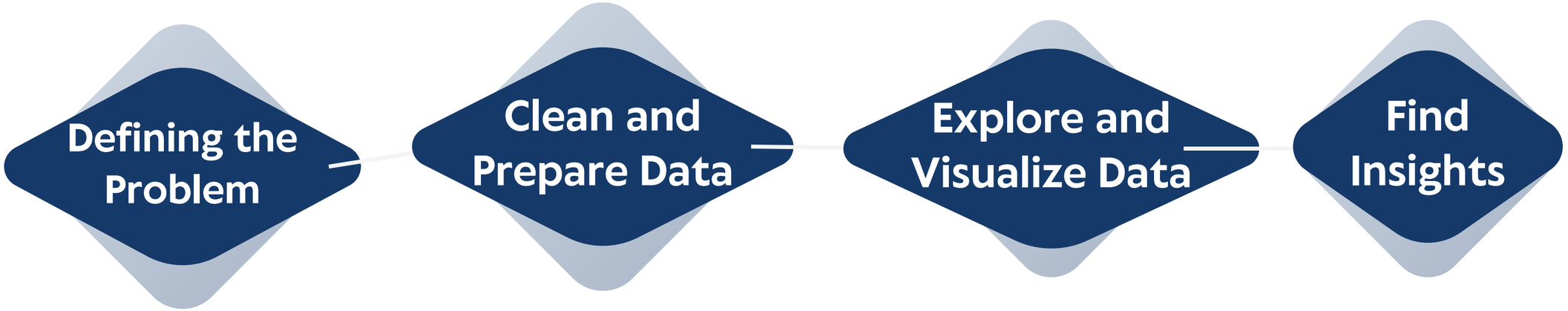
Media & Technology Analysis of YouTube Dataset

-KHUSHEE BINWAL

WORKFLOW

These are the basic steps to be followed to approach the problem.



Clearly articulate the problem you are trying to solve

Handle missing values, remove duplicates, and correct errors

Calculate mean, median, mode, variance, etc. Use charts and graphs to understand data distribution and patterns.

Interpret the results in the context of the problem you are solving

Understanding the Dataset

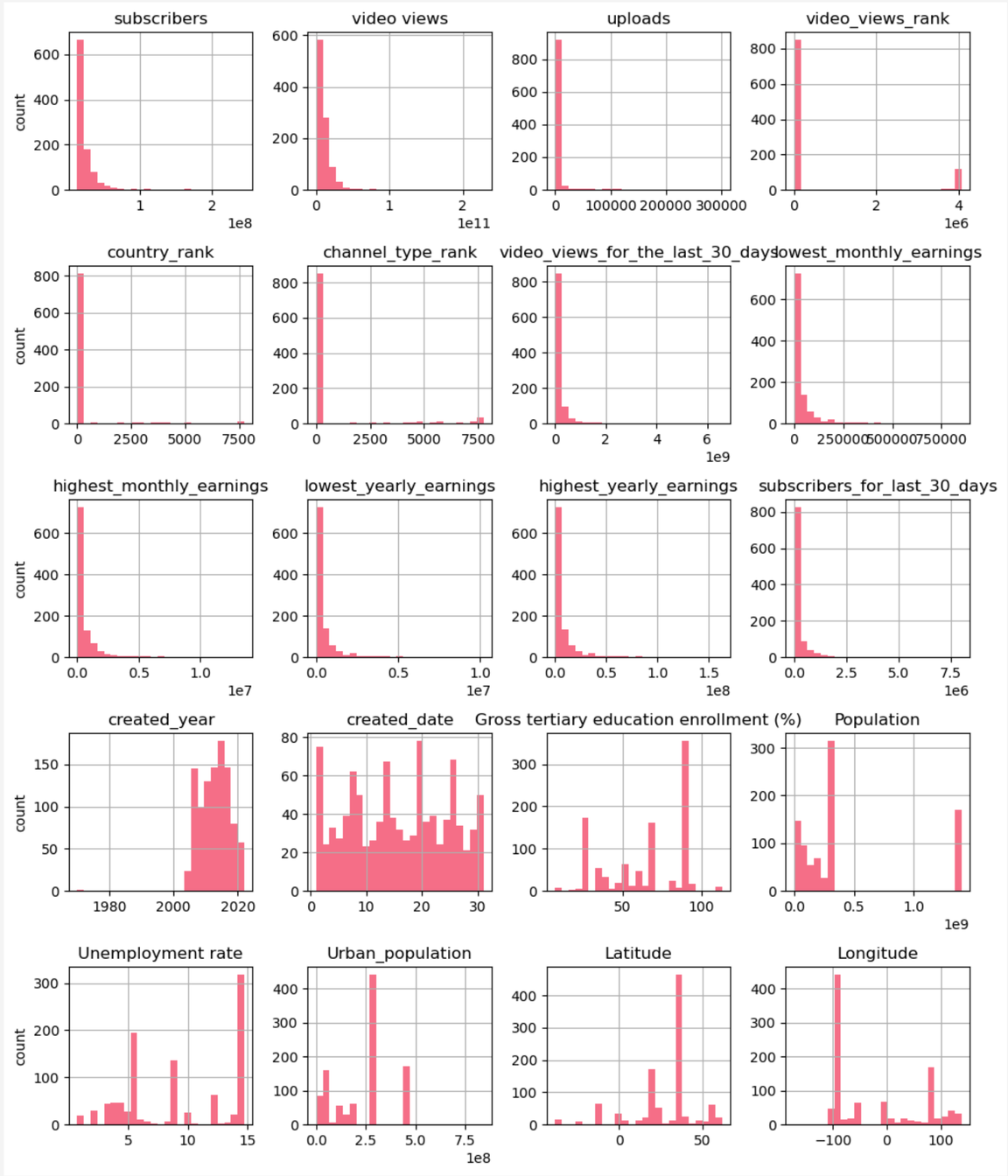
After loading the dataset ,its important to understand the nature of dataset to find insights from the dataset.

DATA CLEANING AND HANDLING MISSING VALUES

Null values were present in many columns . The missing values are handled using appropriate replacement values.

OUTLIERS

Outliers in YouTube data could signal viral content, niche appeal, or trending topics, impacting subscriber growth significantly.

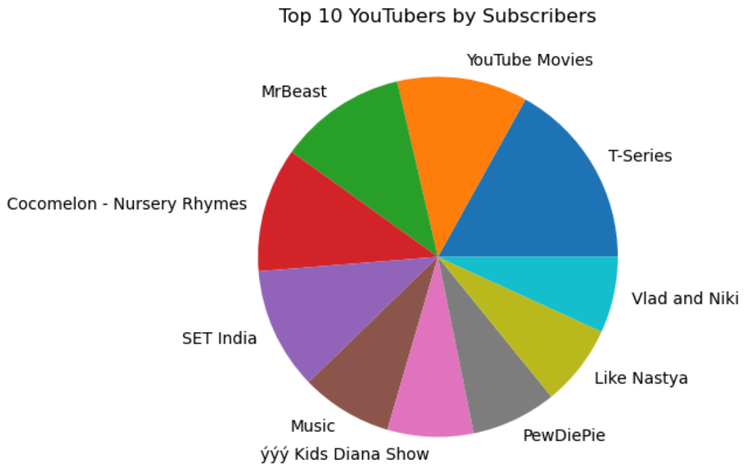


1. What are the top 10 YouTube channels based on the number of subscribers?

rank	
1	T-Series
2	YouTube Movies
3	MrBeast
4	Cocomelon - Nursery Rhymes
5	SET India
6	Music
7	ýýý Kids Diana Show
8	PewDiePie
9	Like Nastya
10	Vlad and Niki

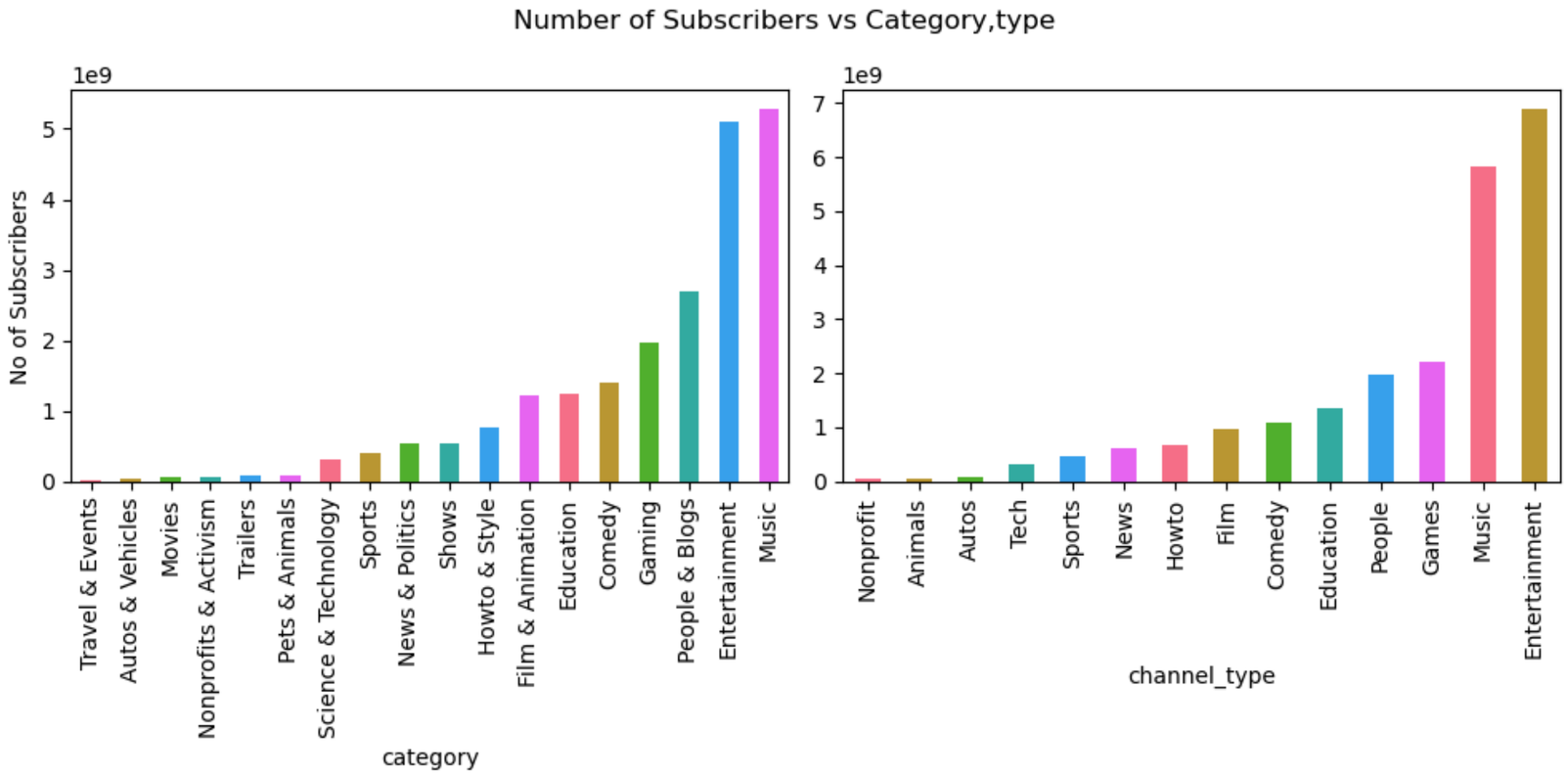
category	
Entertainment	3
People & Blogs	2
Music	1
Film & Animation	1
Education	1
Shows	1
Gaming	1

Country	
United States	6
india	1
India	1
Japan	1
Russia	1



Out of the top 10 youtubers , 3 belong to the Among the top 10 YouTubers, three are from the Entertainment category, and People and Blog also have significant representation. Notably, six out of these top YouTubers are based in the United States, highlighting the country's dominance in producing popular online content creators. This reflects the global influence of U.S. digital media. category followed by People and Blog and most of these (6) belong to United States.

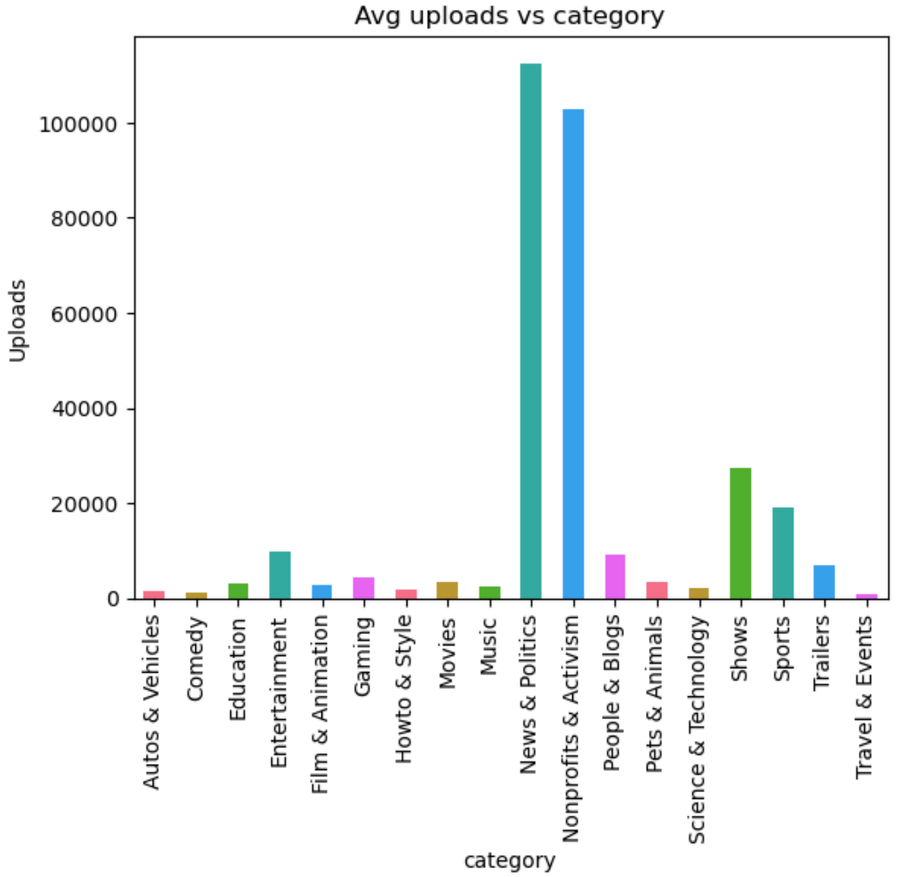
2. Which category has the highest average number of subscribers?



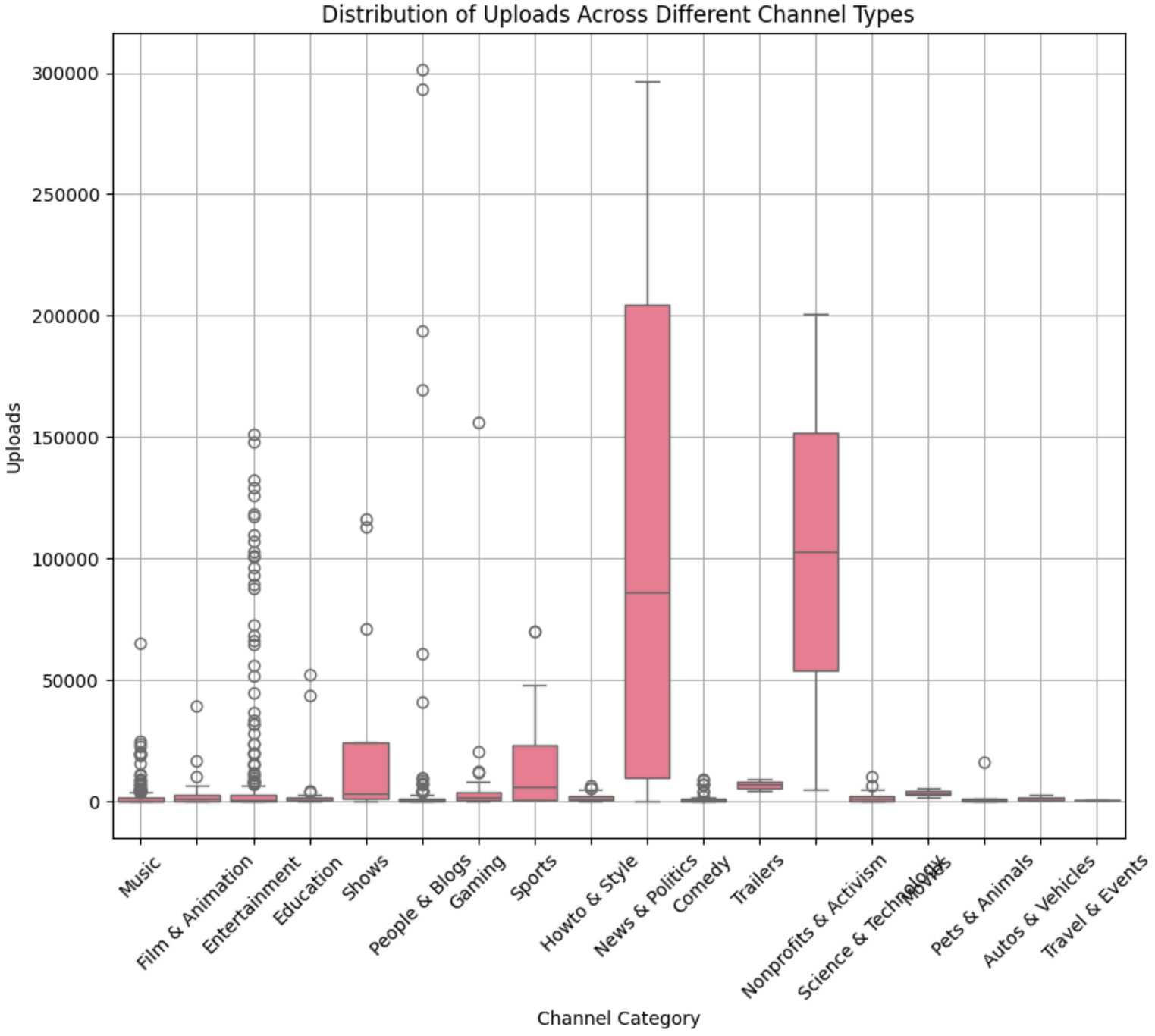
The Entertainment category leads in subscribers, followed by Music, and People and Blogs, with Games trailing. Significant outliers in subscriber counts highlight the vast disparity among YouTube channels, indicating a few with massive followings while many are still growing. This underscores the competitive and varied nature of YouTube popularity.

3. How many videos, on average, are uploaded by YouTube channels in each category?

category	
Entertainment	293
Music	203
People & Blogs	131
Gaming	95
Comedy	70
Education	46
Film & Animation	45
Howto & Style	40
News & Politics	26
Science & Technology	17
Shows	13
Sports	12
Pets & Animals	5
Autos & Vehicles	3
Trailers	2
Nonprofits & Activism	2
Movies	2
Travel & Events	1

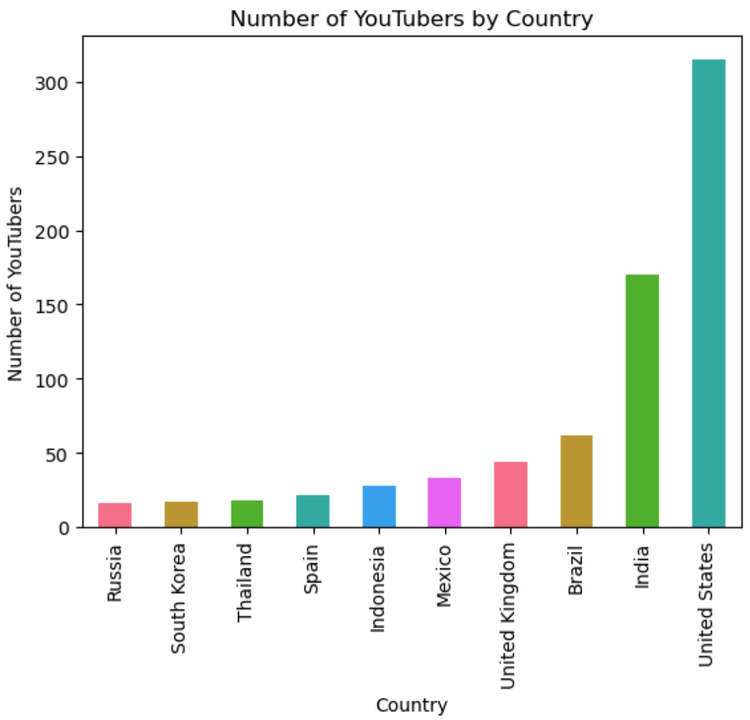


The News and Politics category has the highest number of uploads, followed by Non-Profit Organizations and Activism. Despite the relatively low number of News and Politics channels, their average uploads are high, indicating that channels in this category frequently produce content, likely due to the fast-paced nature of news reporting. This abundance of uploads contributes to significant outliers in the average number of uploads, reflecting the intensive content generation by a few highly active channels.




```
category
News & Politics           112484.384615
Nonprofits & Activism     102912.000000
Shows                      27443.692308
Sports                     19129.833333
Entertainment               9938.399317
People & Blogs              9256.793893
Trailers                    6839.000000
Gaming                      4285.273684
Pets & Animals              3562.800000
Movies                      3553.000000
Education                   3087.086957
Film & Animation            2861.844444
Music                       2325.945813
Science & Technology        2114.058824
Howto & Style               1695.500000
Autos & Vehicles            1550.666667
Comedy                      1202.557143
Travel & Events              766.000000
Name: uploads, dtype: float64
Maximum uploads for: News & Politics ;
with avg uploads of: 112484.38461538461
```

4. What are the top 5 countries with the highest number of YouTube channels?



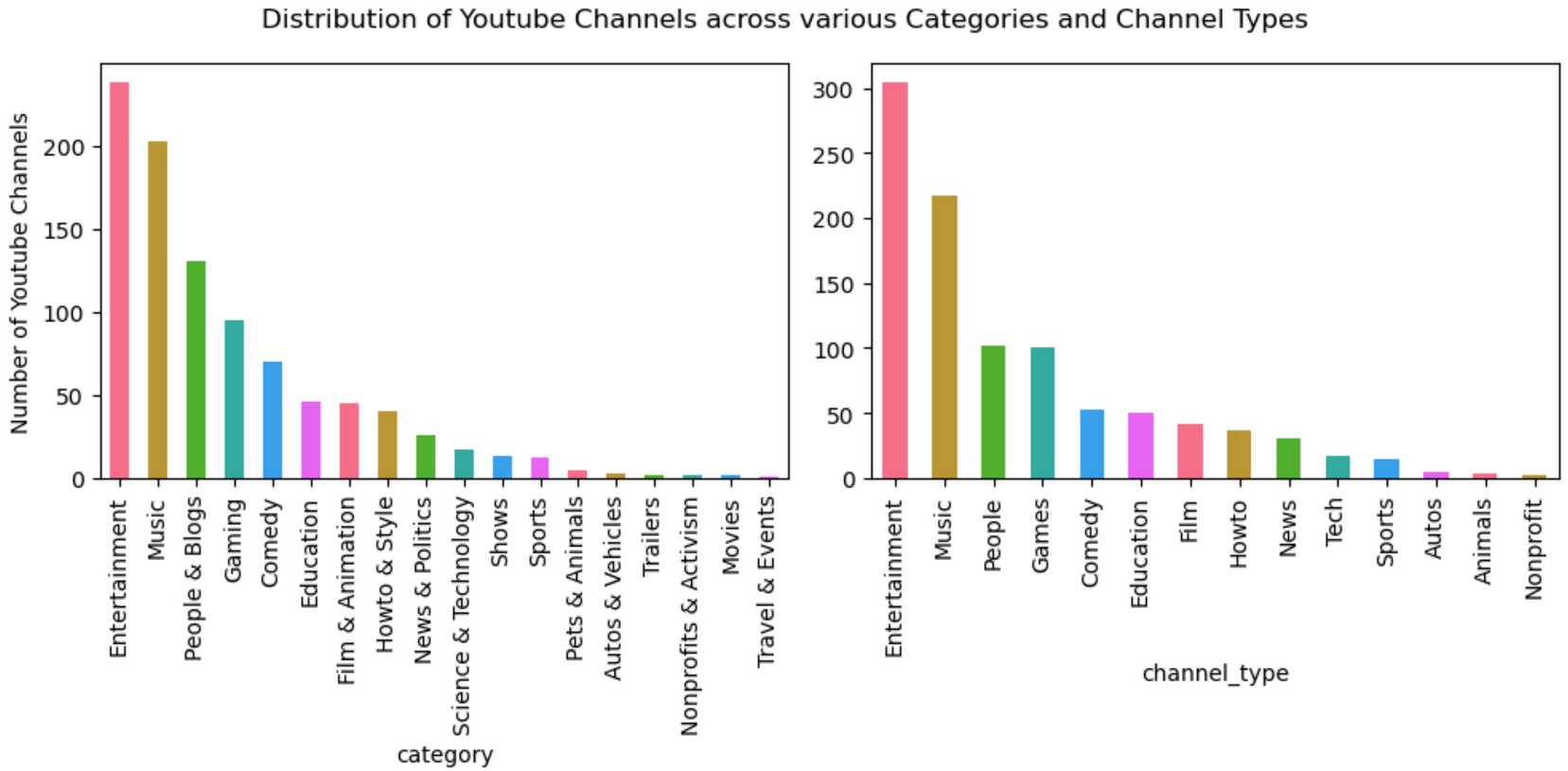
category	
Entertainment	127
Music	95
People & Blogs	66
Gaming	39
Comedy	25
Film & Animation	21
Education	20
Howto & Style	18
Science & Technology	8
Sports	6
Shows	5
News & Politics	5
Pets & Animals	4
Nonprofits & Activism	1

Brazil: category	
Music	15
Entertainment	15
Comedy	8
Film & Animation	7
People & Blogs	6
Gaming	6
Howto & Style	2

India: category	
Entertainment	50
Music	29
Education	16
Comedy	16
People & Blogs	15
News & Politics	11
Film & Animation	11
Howto & Style	7

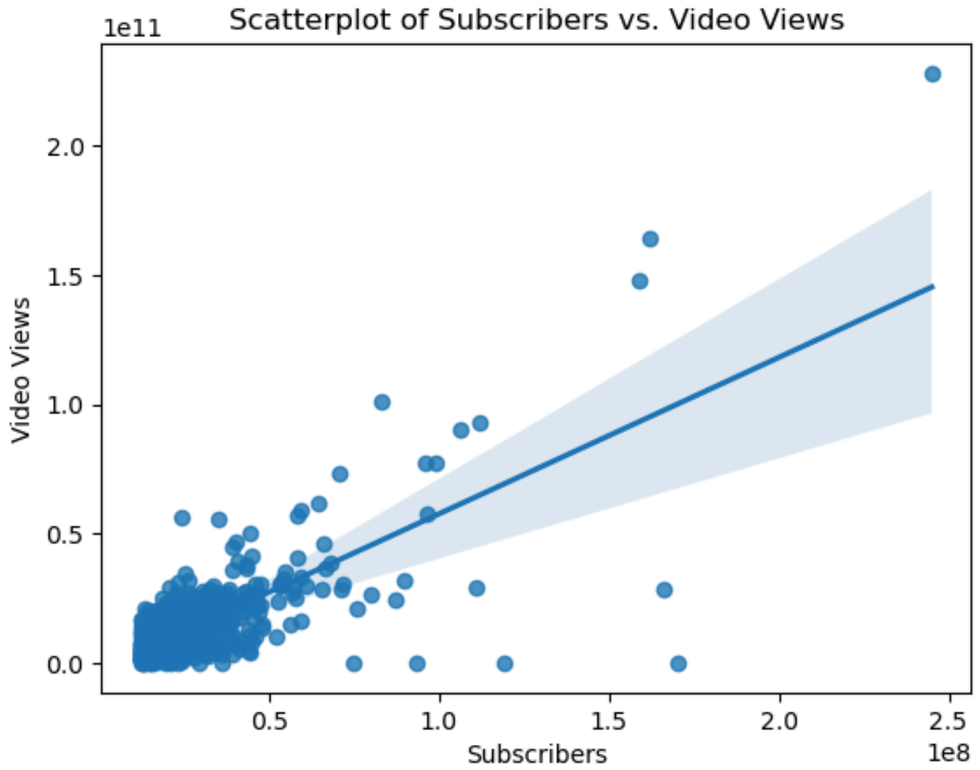
The United States has the most YouTube channels, largely featuring Entertainment, Music, People and Blogs, and Gaming. India and Brazil follow with numerous channels, also emphasizing Entertainment and Music. This reflects global trends in content creation and consumption, highlighting these countries' significant roles in digital media. The popularity of these categories underscores their universal appeal, shaping the diverse landscape of YouTube content worldwide.

5. What is the distribution of channel types across different categories?



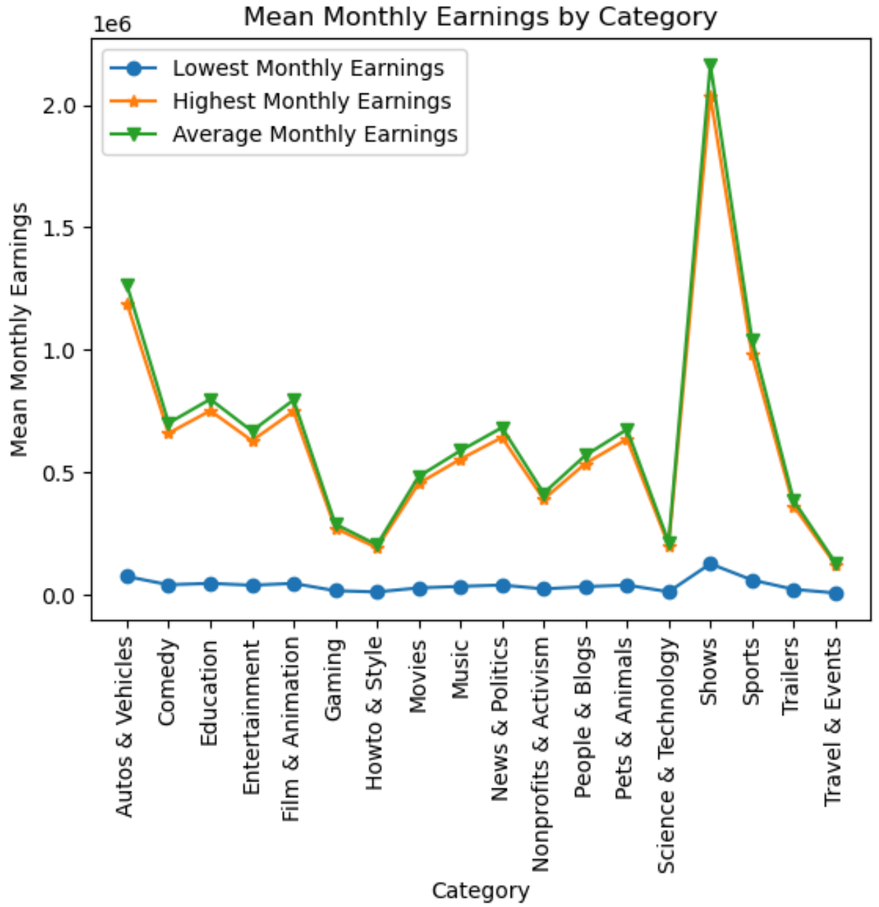
The prevalence of Entertainment, Music, People and Blogs, and Gaming channels suggests strong viewer interest in diverse forms of entertainment and personal expression on YouTube. These categories dominate due to their ability to engage large audiences globally through varied and appealing content.

6. Is there a correlation between the number of subscribers and total video views for YouTube channels?



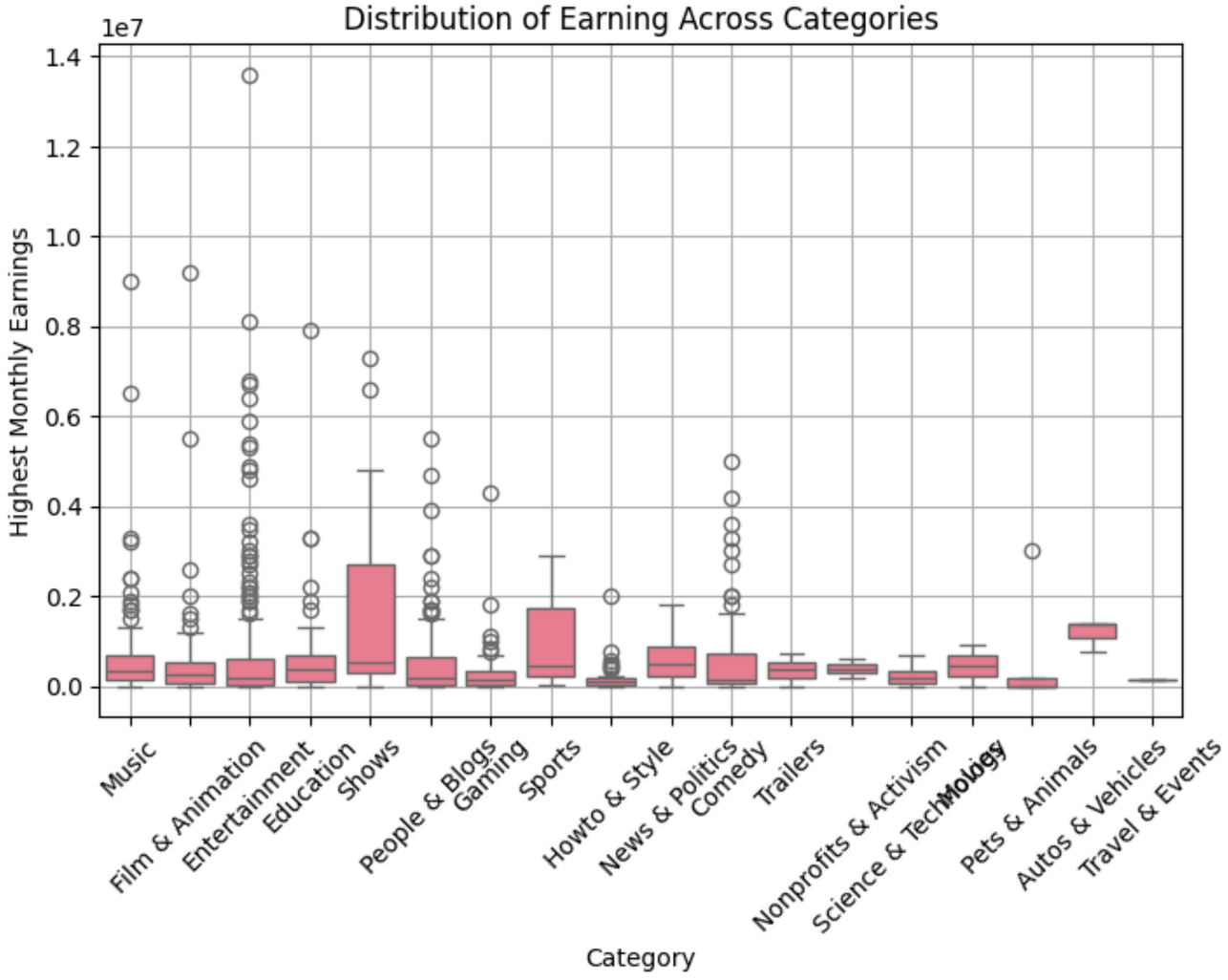
The scatter plot confirms a positive correlation (correlation factor: 0.75084) between subscribers and video views on YouTube channels. Channels with fewer subscribers generally receive fewer to moderate views, as indicated by a dense cluster in the lower left. However, as subscriber counts increase, so do video views, suggesting that popular channels attract larger audiences. This trend underscores how subscriber base impacts viewer engagement and channel visibility on the platform.

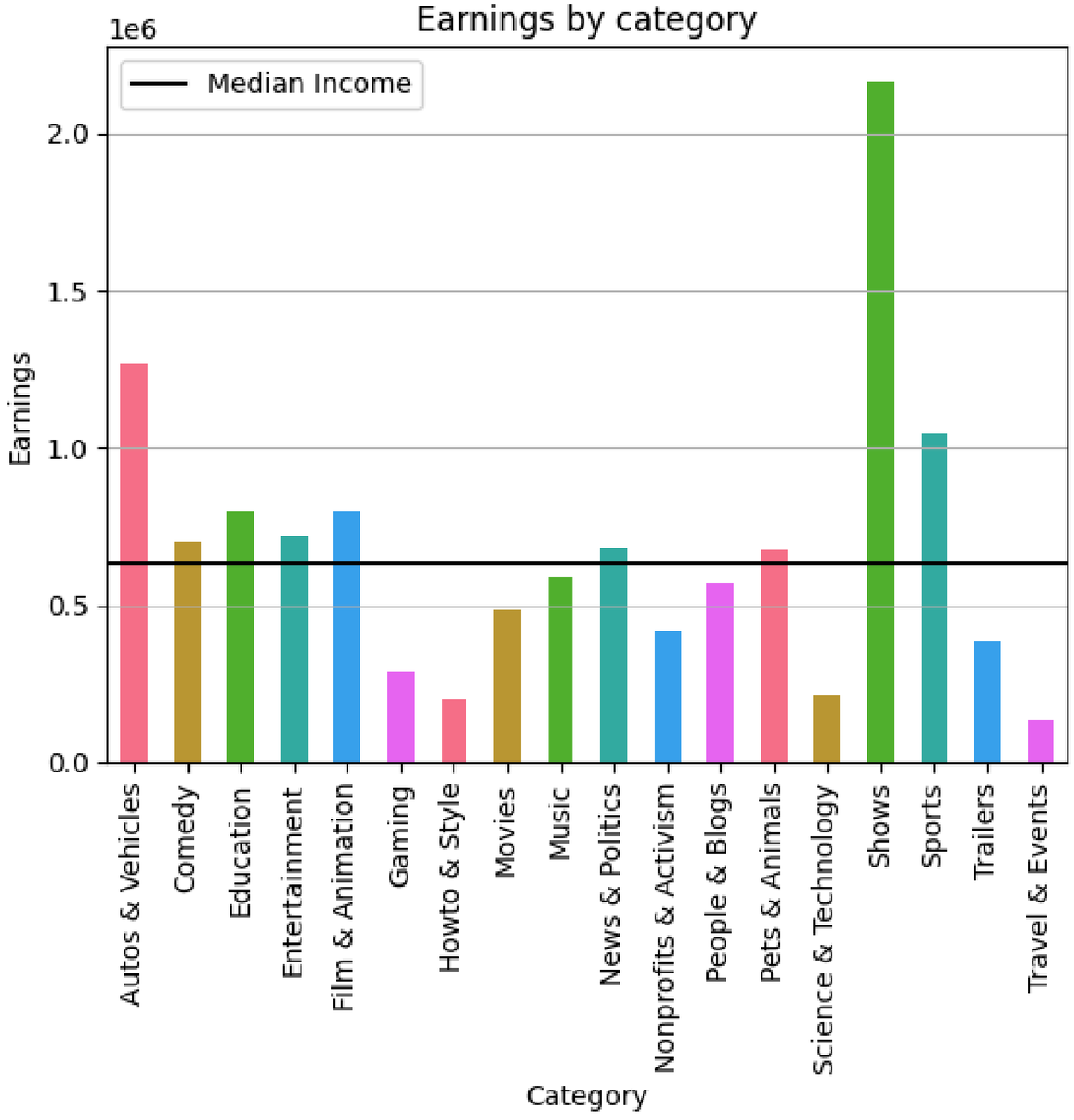
7. How do the monthly earnings vary throughout different categories?



Monthly Mean Income by category	
Travel & Events	1.318000e+05
Howto & Style	2.040664e+05
Science & Technology	2.146786e+05
Gaming	2.889859e+05
Trailers	3.845000e+05
Nonprofits & Activism	4.148000e+05
Movies	4.831000e+05
People & Blogs	5.692353e+05
Music	5.882452e+05
Pets & Animals	6.755324e+05
News & Politics	6.828246e+05
Comedy	7.010958e+05
Entertainment	7.164127e+05
Film & Animation	7.957866e+05
Education	7.986676e+05
Sports	1.042142e+06
Autos & Vehicles	1.265867e+06
Shows	2.164623e+06

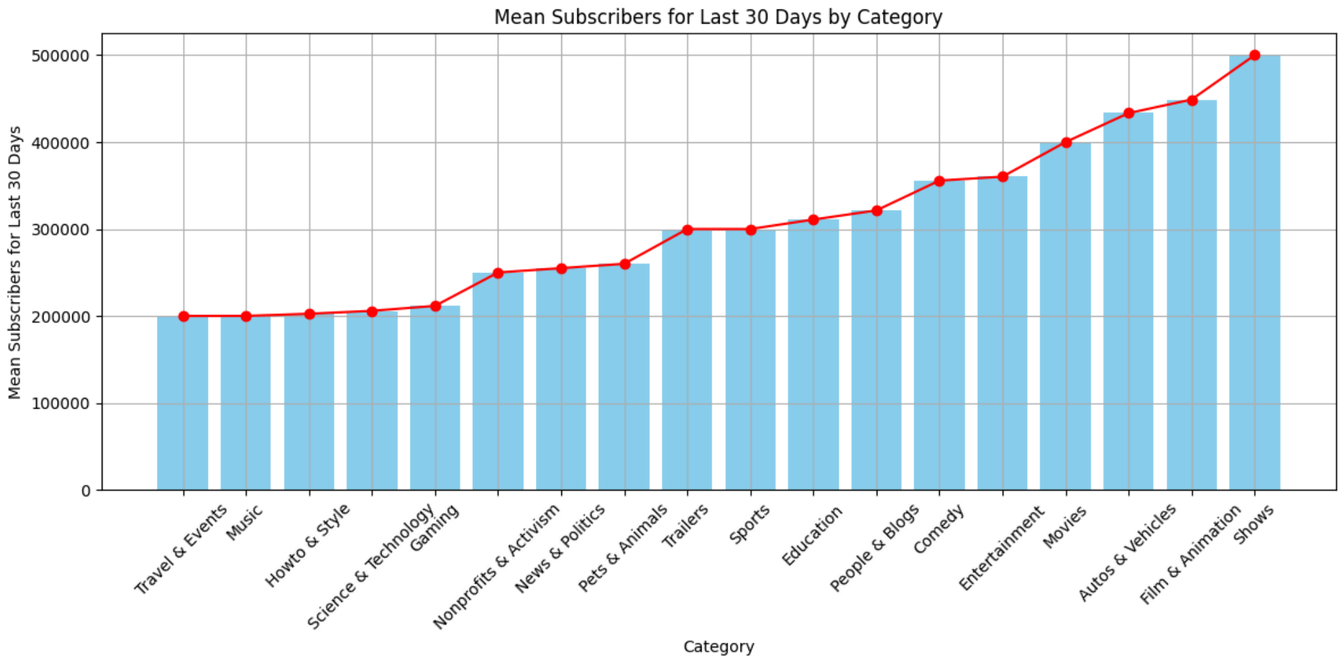
There exists a significant disparity between the lowest and highest monthly earnings among YouTube channel categories. On average, channels tend to earn closer to the highest monthly earnings, highlighting the wide gap between extremes. Categories like Automobiles, Sports, Shows, Education, and Comedy generally earn above the median income, indicating stronger monetization potential. Conversely, Gaming, Movies, and Music categories typically yield lower earnings, underscoring varying financial success across different content genres.


Distribution of Earning Across Categories
1e7
Highest Monthly Earnings
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Music
Film & Animation
Entertainment
Education
Shows
People & Blogs
Gaming
Sports
Howto & Style
News & Politics
Comedy
Trailers
Nonprofits & Activism
Science & Technology
Movies
Pets & Animals
Autos & Vehicles
Travel & Events
Category


Earnings by category
1e6
Median Income
Earnings
2.0
1.5
1.0
0.5
0.0
Autos & Vehicles
Comedy
Education
Entertainment
Film & Animation
Gaming
Howto & Style
Movies
Music
News & Politics
Nonprofits & Activism
People & Blogs
Pets & Animals
Science & Technology
Shows
Sports
Trailers
Travel & Events
Category

8. What is the overall trend in subscribers gained in the last 30 days across all channels?

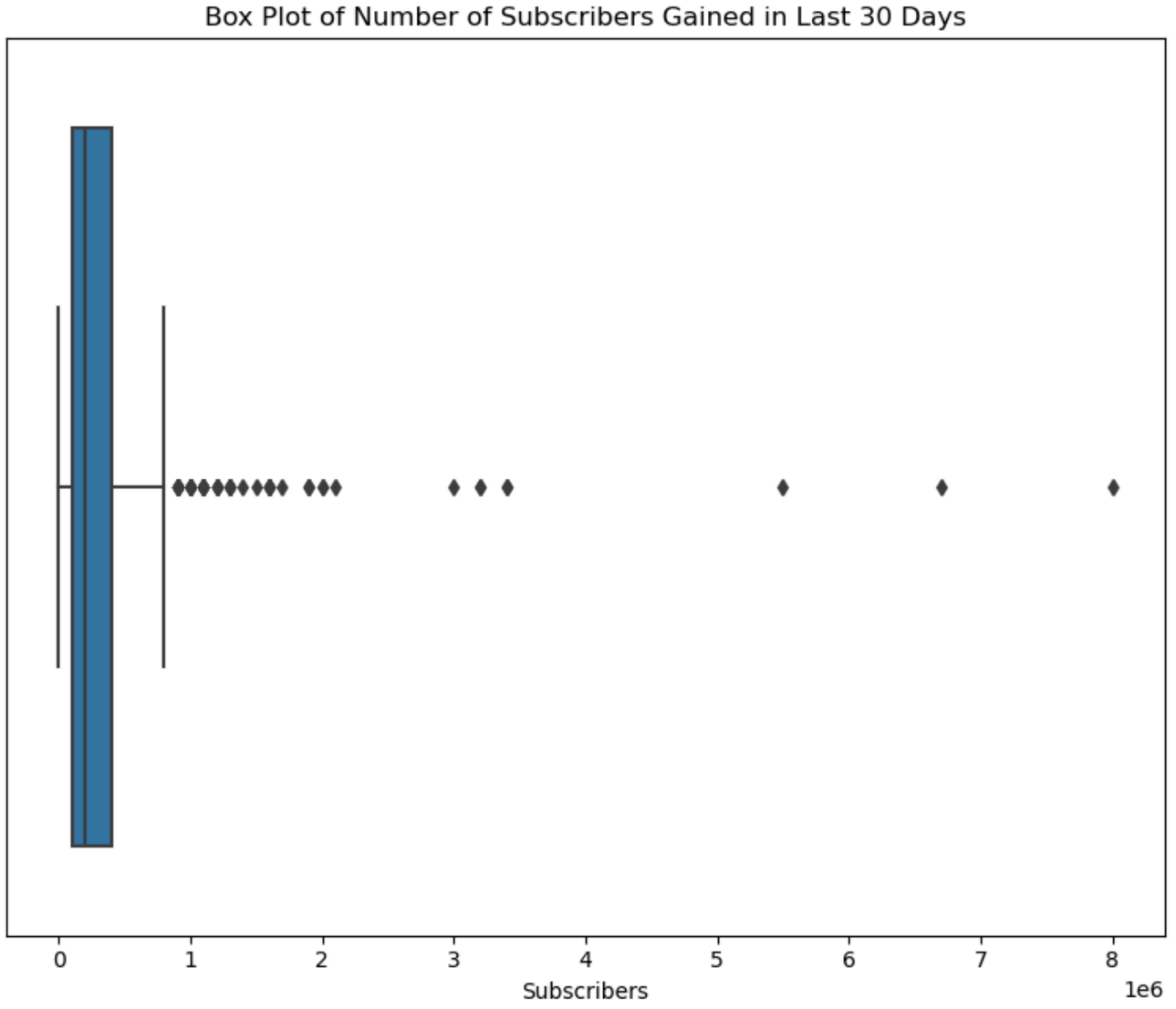
Briefly elaborate on what you want to discuss.



TV shows likely attract the most subscribers, followed by films, animations, automotive content, and general movies and entertainment. Each category can experience outliers, driven by viral content or niche appeal. For instance, breakout series or viral videos of unique cars can significantly spike subscriber numbers. Factors like content quality, engagement strategies, and timely capitalization on trends are crucial. Channels benefitting from these factors tend to see sustained growth, despite occasional outliers influencing subscriber metrics unpredictably.

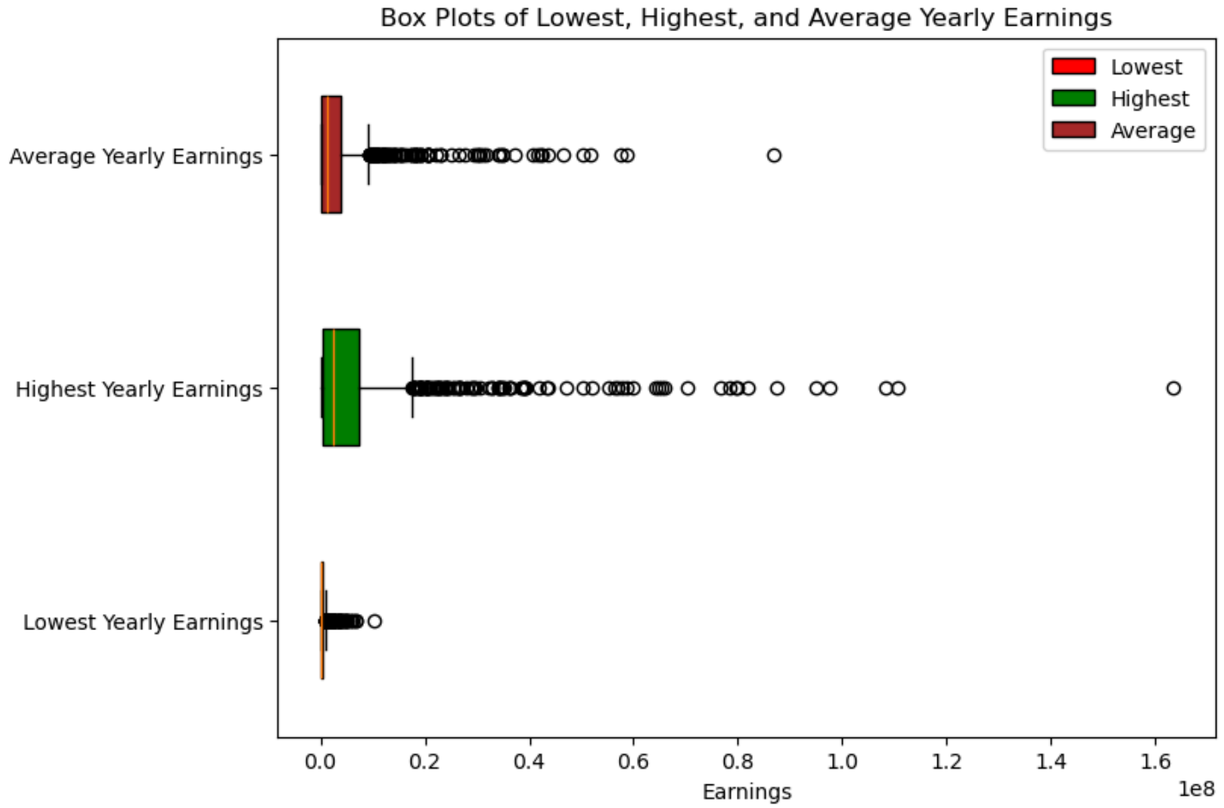
category	
Travel & Events	200000.000000
Music	200055.290640
Howto & Style	202500.075000
Science & Technology	205884.823529
Gaming	211579.357895
Nonprofits & Activism	250000.000000
News & Politics	255000.000000
Pets & Animals	260002.000000
Trailers	300000.000000
Sports	300000.000000
Education	310869.565217
People & Blogs	321425.244275
Comedy	355717.942857
Entertainment	360217.153584
Movies	400000.000000
Autos & Vehicles	433333.333333
Film & Animation	448954.777778
Shows	500000.153846

count	6.660000e+02
mean	3.495419e+05
std	6.131554e+05
min	1.000000e+00
25%	1.000000e+05
50%	2.000000e+05
75%	4.000000e+05
max	8.000000e+06



9. Are there any outliers in terms of yearly earnings from YouTube channels?

Briefly elaborate on what you want to discuss.



Upper Fence: 9282106.25 Lower_Fence: -5126643.75

Number of ouliers on right Side 101

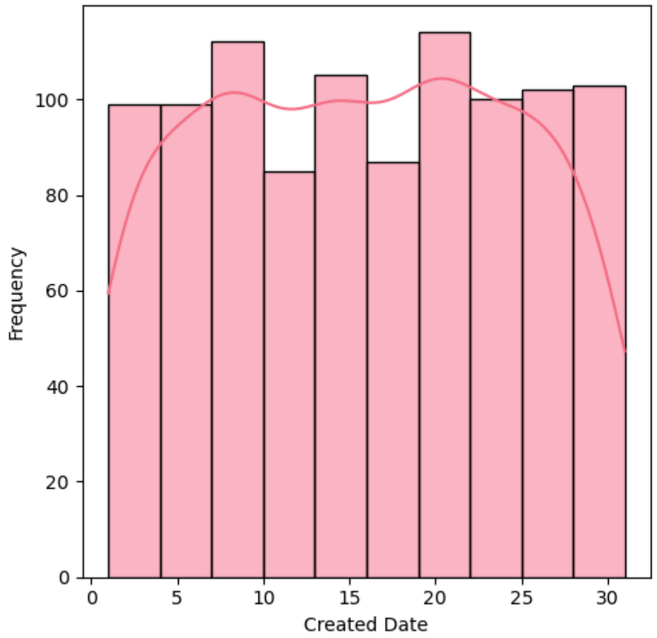
Outliers exist in earnings from youtube, these Outliers represent data points that are unusually far from other observations. They can indicate extremes in the dataset, such as exceptionally high or low values. the calculation of outliers here is made using the:

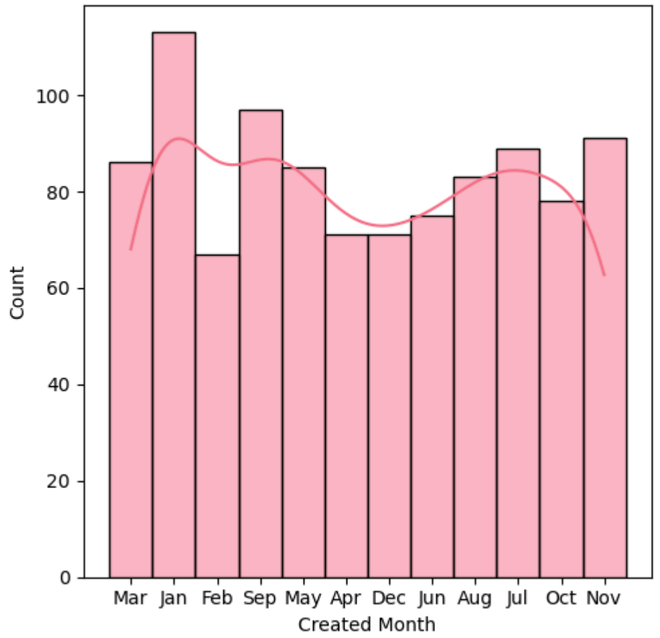
1) Calculate the IQR as IQR=Q3−Q1IQR = Q3 - Q1IQR=Q3−Q1.

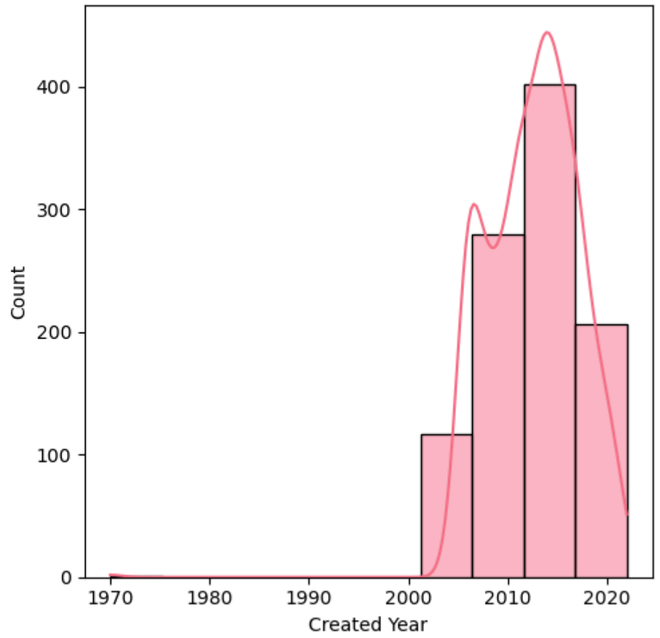
2)Define a lower bound Q1−1.5×IQR and an upper bound Q3+1.5×IQR .

Any value outside these are outliers.

10. What is the distribution of channel creation dates? Is there any trend over time?



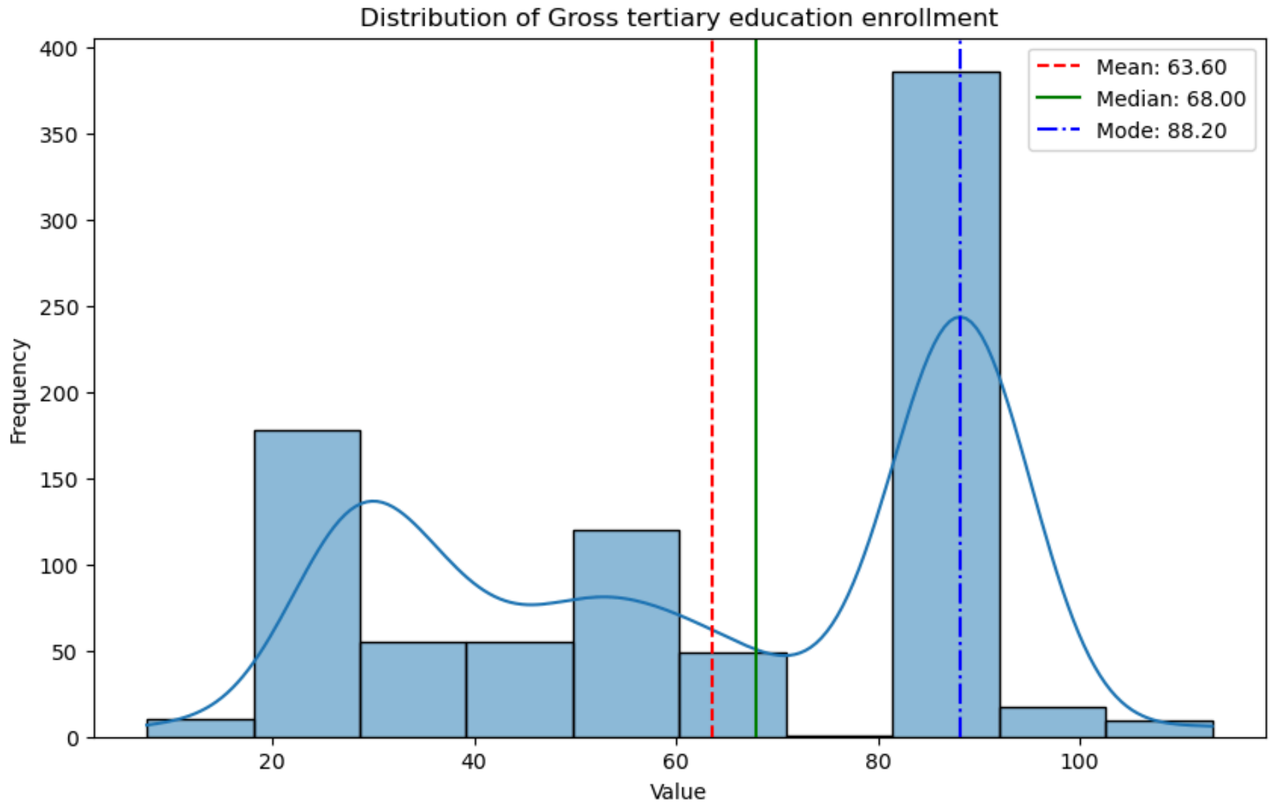


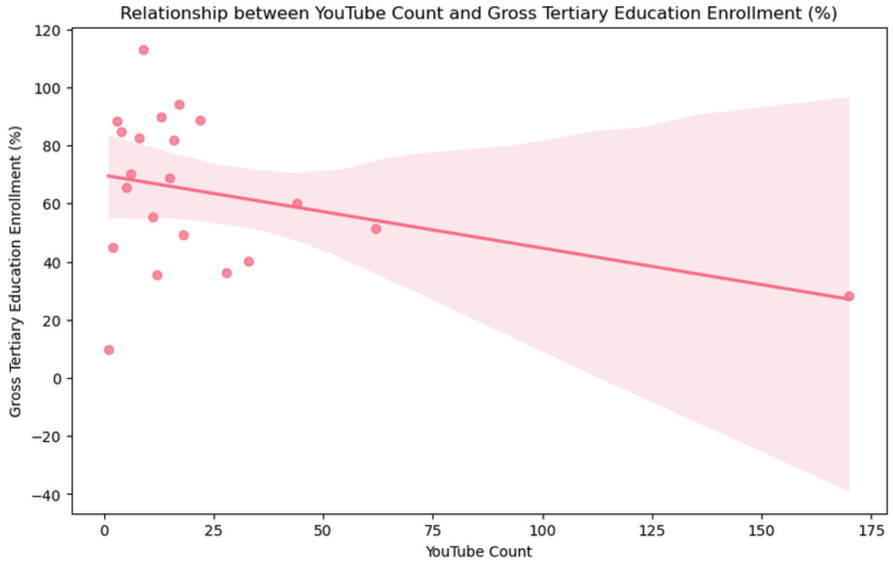


Based on the data provided, the trend of YouTube channel creation shows peaks around 2005 and again in 2015, suggesting a roughly normal distribution over time. This pattern implies initial enthusiasm during YouTube's early days, followed by renewed interest, possibly due to technological advancements, cultural shifts, or changes in content consumption habits. This reflects the dynamic evolution of digital media platforms and the diverse factors influencing content creation trends over time.

11. Is there a relationship between gross tertiary education enrollment and the number of YouTube channels in a country?

Briefly elaborate on what you want to discuss.



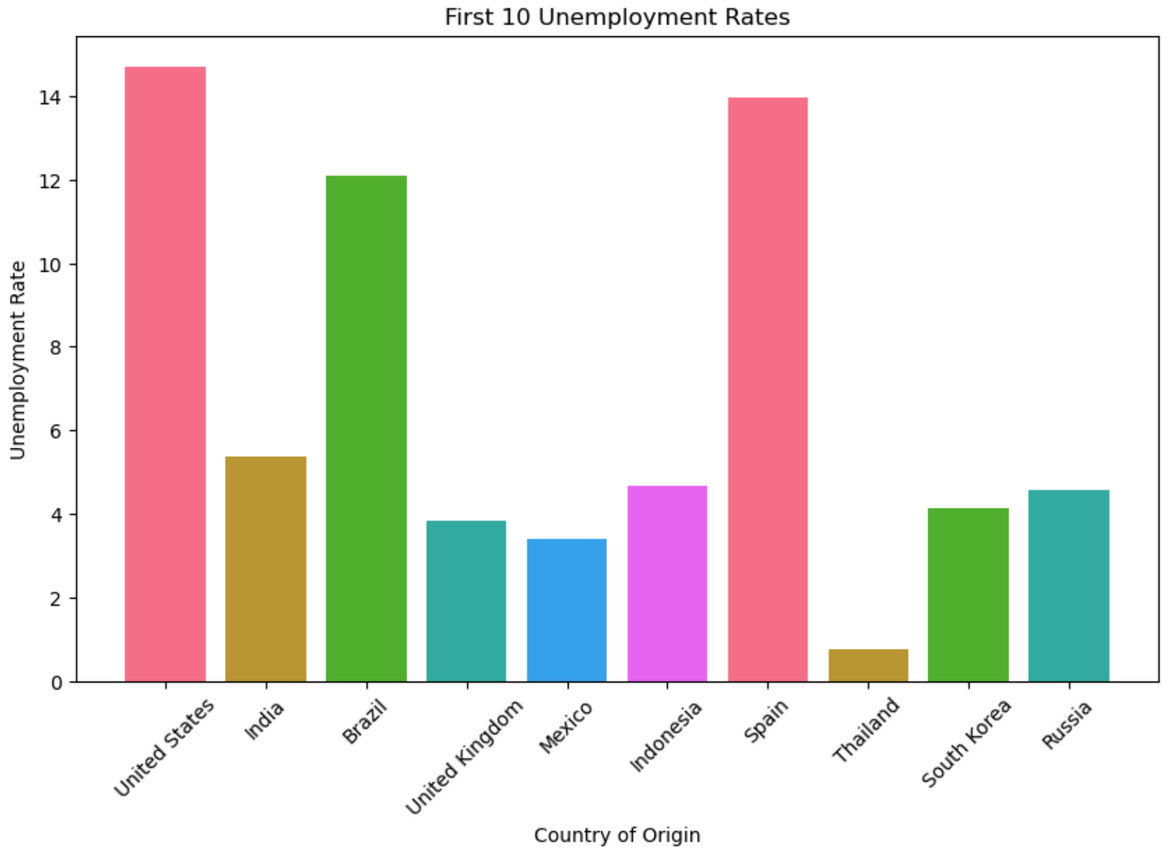



```
Correlation coefficient: -0.3530784655419943
```

A negative correlation of 0.353 suggests that as gross tertiary education enrollment increases in a country, the number of YouTube channels tends to decrease moderately. This relationship implies that higher levels of tertiary education might lead to fewer individuals pursuing YouTube as a platform for content creation, possibly due to career priorities or alternative educational opportunities. However, this correlation does not imply causation and could be influenced by various cultural, economic, or technological factors specific to each country's context.

12. How does the unemployment rate vary among the top 10 countries with the highest number of YouTube channels?

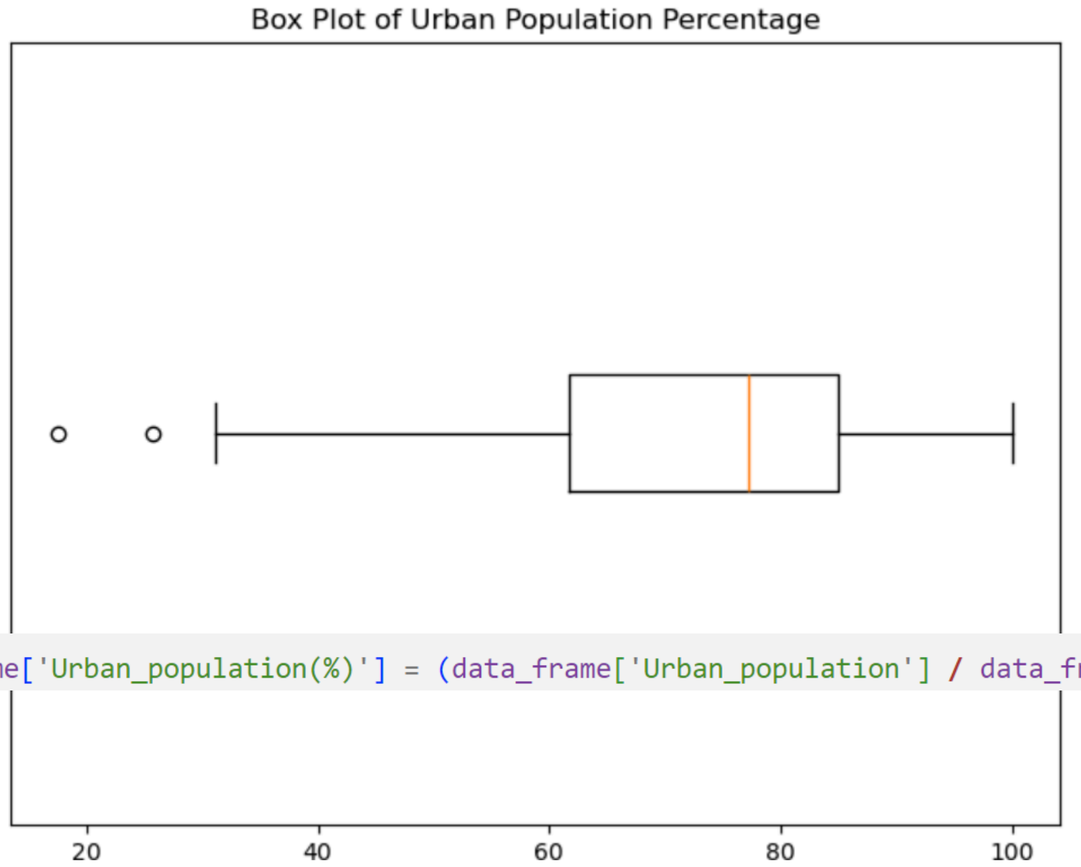
count	10.00000
mean	6.75500
std	4.90686
min	0.75000
25%	3.92500
50%	4.64000
75%	10.40000
max	14.70000



The unemployment rates vary significantly among the top 10 countries with the highest number of YouTube channels. For instance, the United States shows a mean rate of 6%, with peaks at approximately 14%. Similarly, Spain and Brazil exhibit rates nearing 14% and 12%, respectively, contrasting with India's lower rate of around 5%. It's crucial to note that these figures may not represent the broader population accurately due to potential sampling biases in data collection methods, possibly skewing the representation of unemployment rates across different countries.

13. What is the average urban population percentage in countries with YouTube channels?

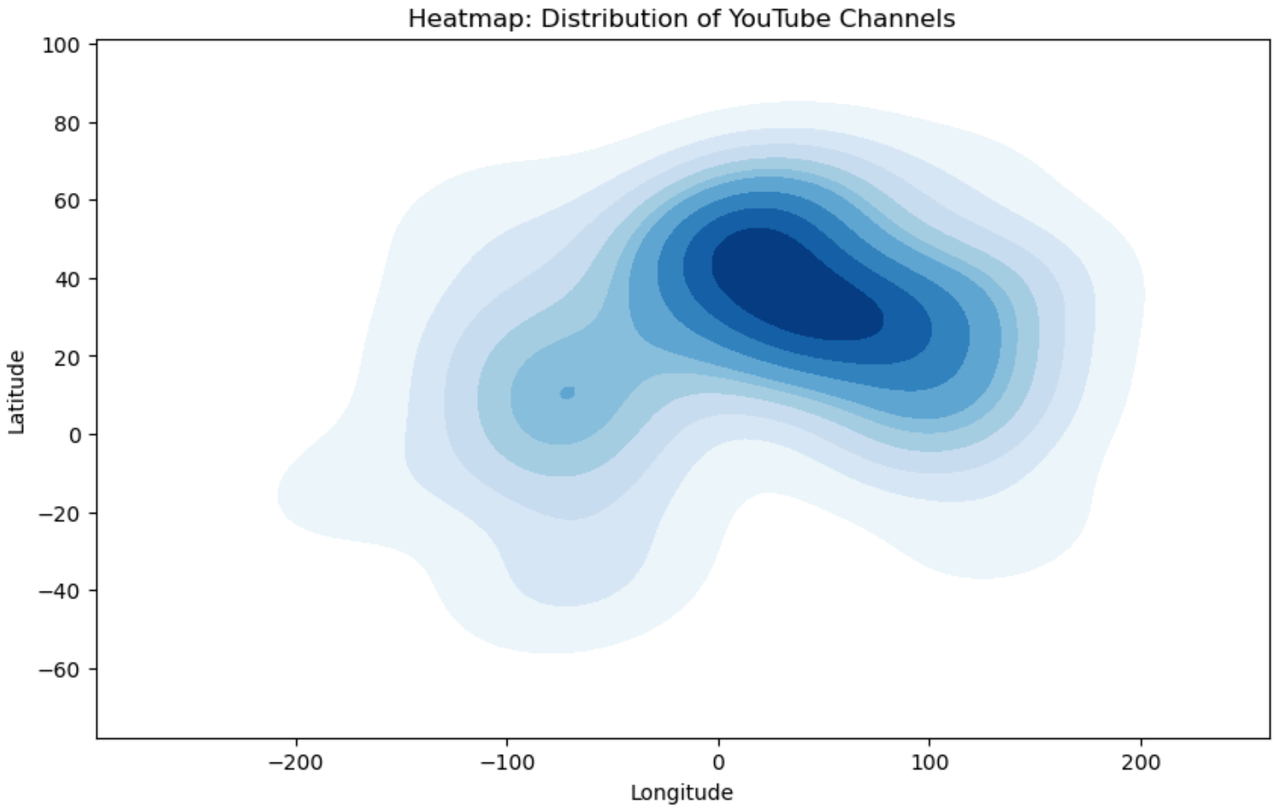
	Urban_population(%)
count	48.000000
mean	70.789476
std	20.649655
min	17.573800
25%	61.760847
50%	77.242498
75%	84.946002
max	100.000000




```
data_frame['Urban_population(%)'] = (data_frame['Urban_population'] / data_frame['Population']) * 100
```

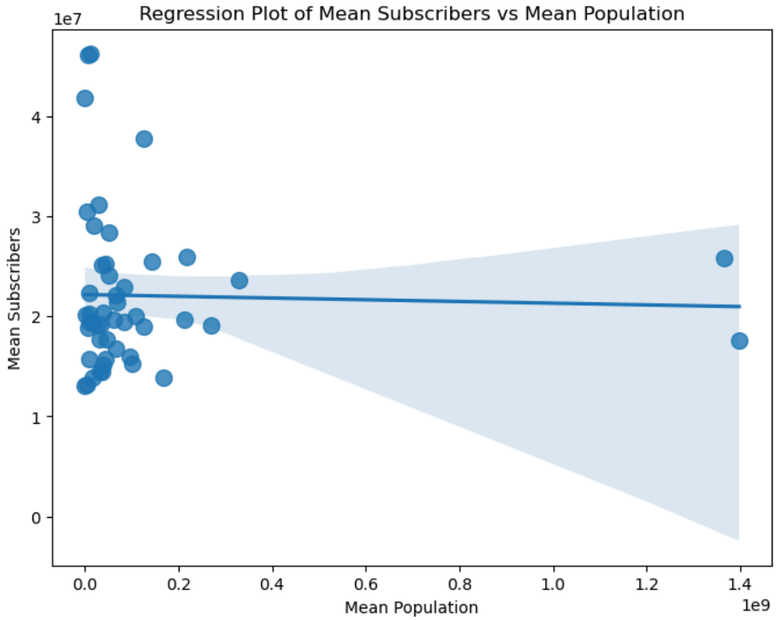
The average urban population rate of 70% with a standard deviation of 20.64 suggests variability around this mean in the dataset. Outliers on the left side of a box plot indicate urban populations significantly lower than the average, possibly due to rural areas or demographic anomalies

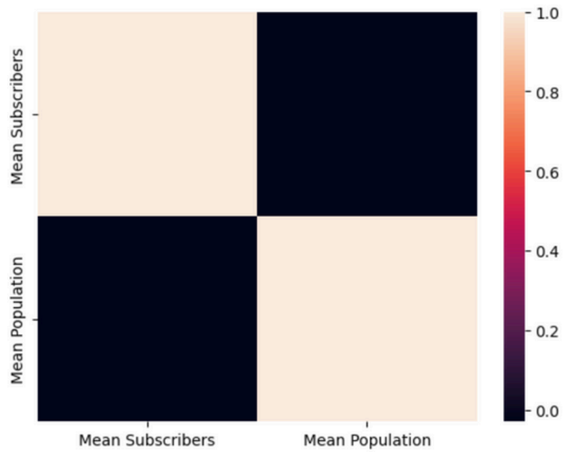
14. Are there any patterns in the distribution of YouTube channels based on latitude and longitude coordinates?



Countries like India, Brazil, and the United States exhibit high density of YouTube channels within the longitude range of -24 and latitude range of 20-30. These nations possess robust digital infrastructures, large populations with internet access, and vibrant cultural scenes that foster prolific content creation. The concentration suggests these regions are pivotal for global digital content trends, influencing viewer demographics and advertising strategies tailored to diverse and expansive online audiences.

15.What is the correlation between the number of subscribers and the population of a country

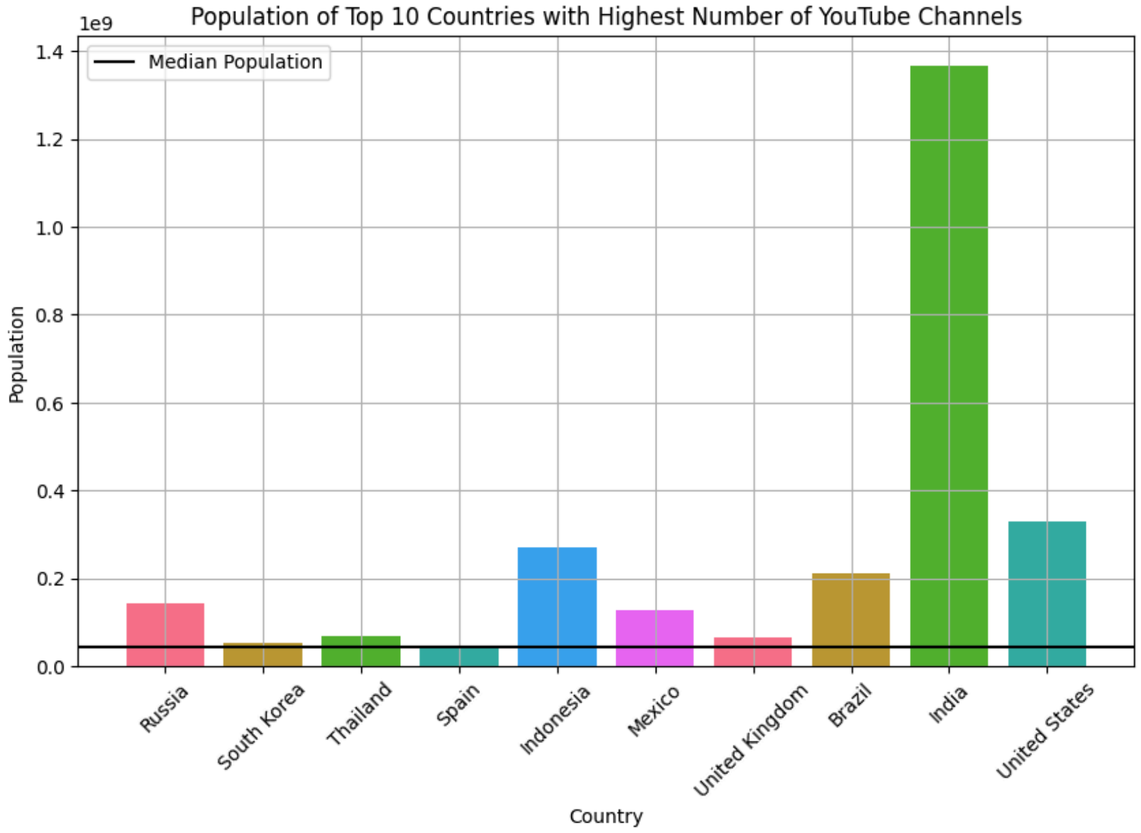





```
Missing values in merged_data:
 Mean Subscribers    0
Mean Population      1
dtype: int64
Correlation Matrix:
                    Mean Subscribers  Mean Population
Mean Subscribers            1.000000        -0.030011
Mean Population            -0.030011         1.000000
```

There is a weak negative correlation of -0.03 between the number of subscribers to a platform like YouTube and a country's population. This suggests that as population size increases, there tends to be a slight decrease in the average number of subscribers per capita. Factors such as internet penetration rates, cultural preferences for digital content, and economic conditions likely influence this relationship across different countries.

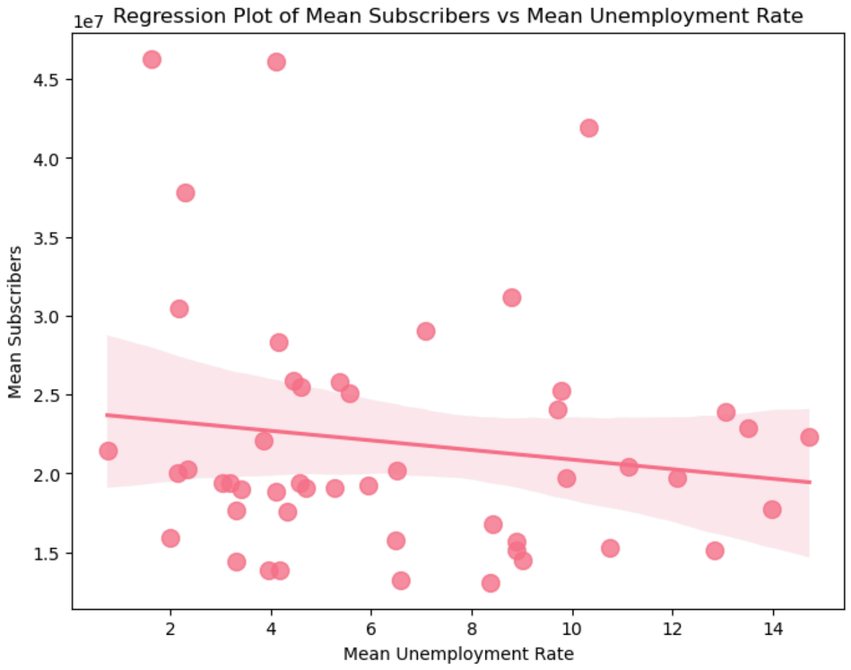
16. How do the top 10 countries with the highest number of YouTube channels compare in terms of their total population?



	Population	No of channels
count	1.000000e+01	10.000000
mean	2.683054e+08	84.900000
std	3.977828e+08	132.952331
min	4.707678e+07	16.000000
25%	6.753220e+07	19.000000
50%	1.351938e+08	30.500000
75%	2.557928e+08	57.500000
max	1.366418e+09	440.000000

The top 10 countries with the most YouTube channels, including India, the United States, and Indonesia, also boast substantial populations exceeding global averages. This correlation (although weak negative) underscores how population size influences digital engagement and content creation trends on platforms like YouTube, reflecting regional demographics and cultural preferences.

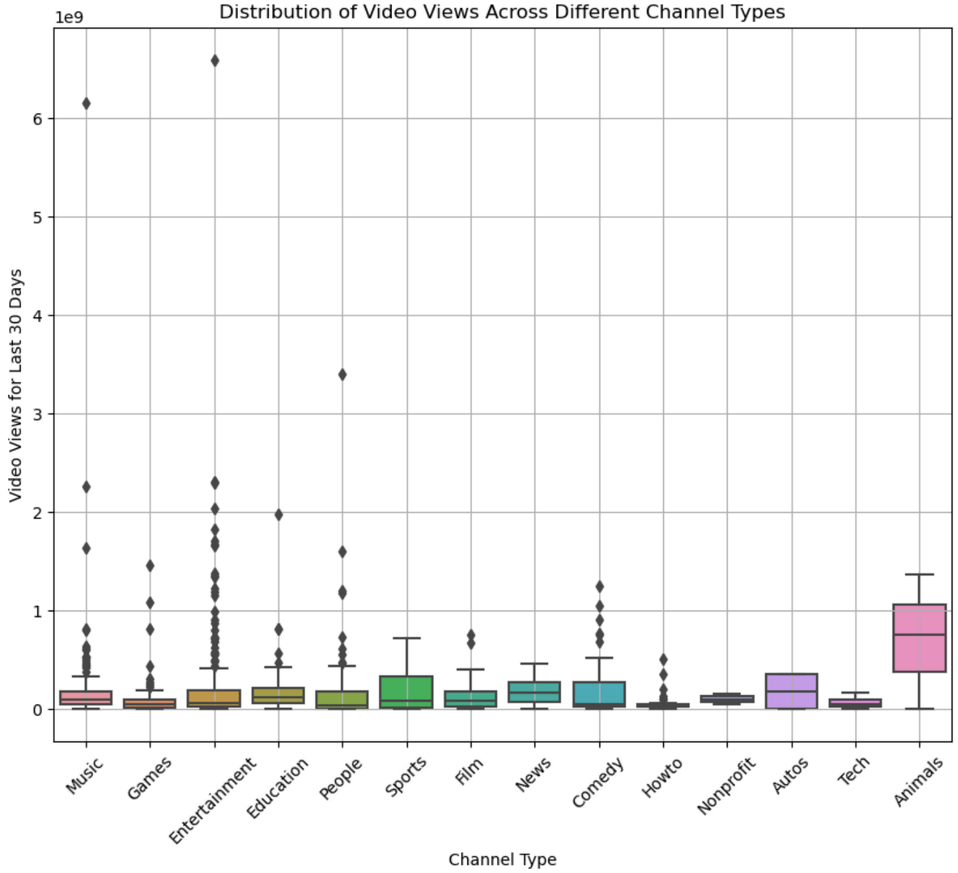
17. Is there a correlation between the number of subscribers gained in the last 30 days and the unemployment rate in a country?

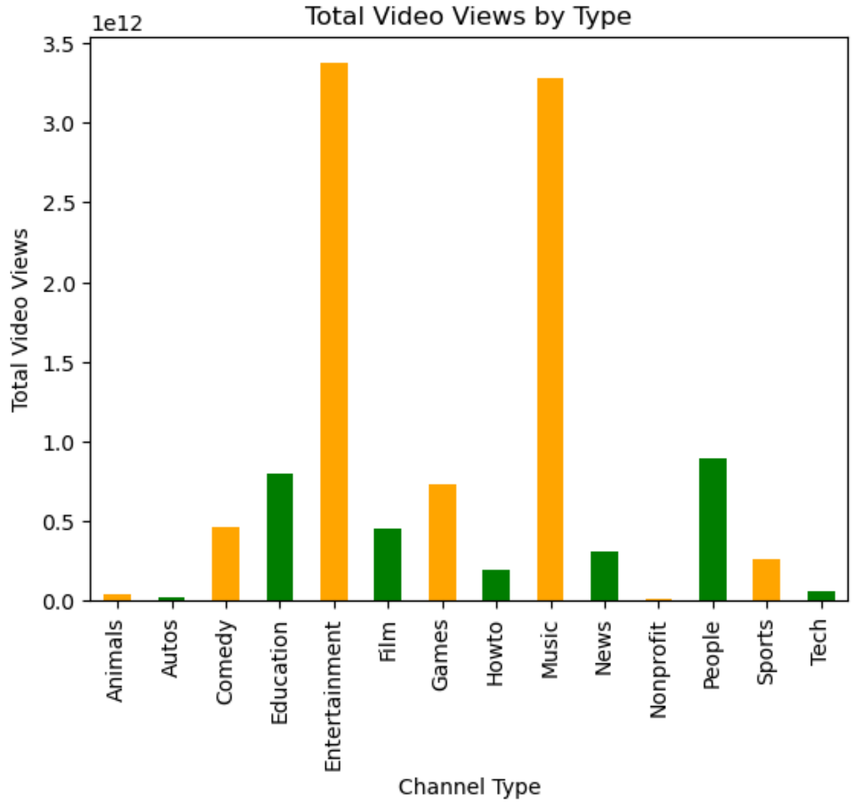


Mean Subscribers	1.000000	-0.140944
Mean UR	-0.140944	1.000000

There is a weak negative correlation of -0.14 between the number of subscribers gained in the last 30 days and the unemployment rate in countries. This suggests that as unemployment increases slightly, there tends to be a modest decrease in new subscriber growth, possibly reflecting economic conditions impacting digital engagement.This correlation suggests that economic factors, such as job market conditions, could influence individuals' ability or inclination to engage with online platforms actively. However, the effect is subtle, indicating that other factors like internet access and content trends also play significant roles in subscriber acquisition dynamics.

18. How does the distribution of video views for the last 30 days vary across different channel types?

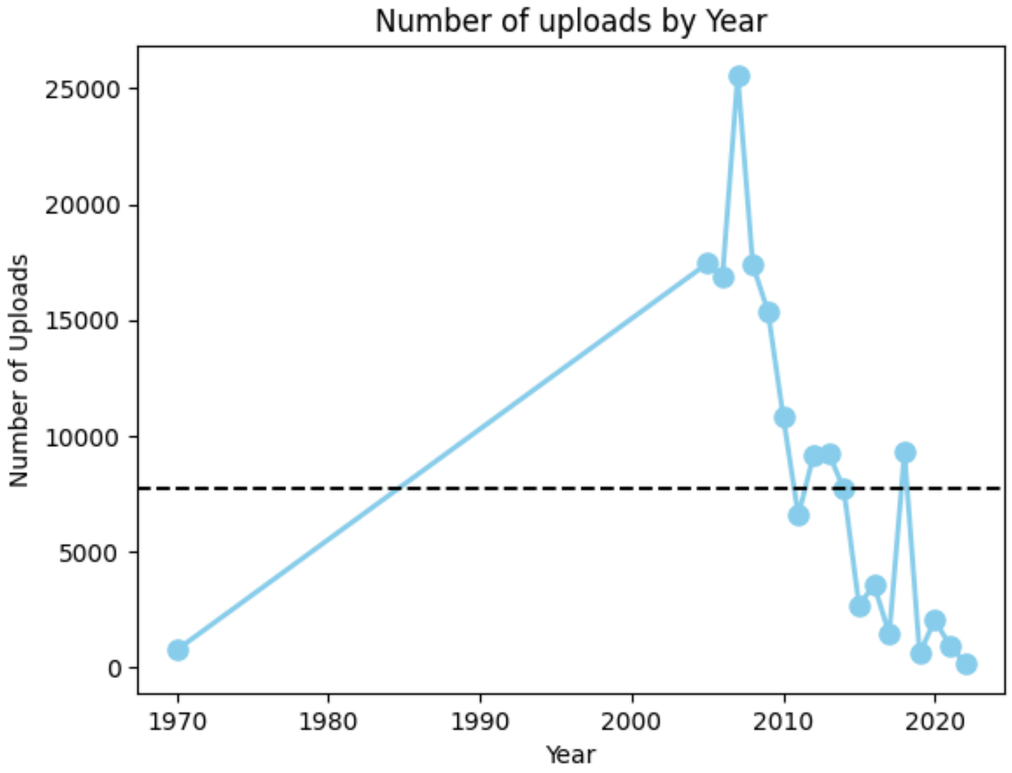


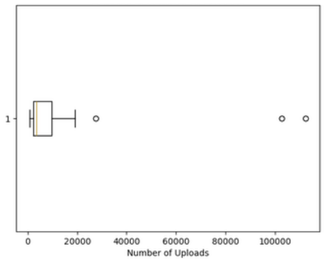


count	1.400000e+01
mean	7.993467e+11
std	1.171431e+12
min	1.086291e+10
25%	9.413713e+10
50%	3.852978e+11
75%	7.791757e+11
max	3.662124e+12

In YouTube subscriber metrics, categories like Entertainment, Comedy, and Music often exhibit higher subscriber gains compared to others. Outliers within these categories suggest exceptional performances, such as viral videos or popular releases, significantly boosting subscriber numbers. Entertainment and Music categories, in particular, benefit from diverse content appeal and celebrity influence, driving subscriber spikes. Understanding these outliers is crucial for content creators and marketers to capitalize on trends and tailor strategies that resonate widely and potentially lead to sustained audience growth.

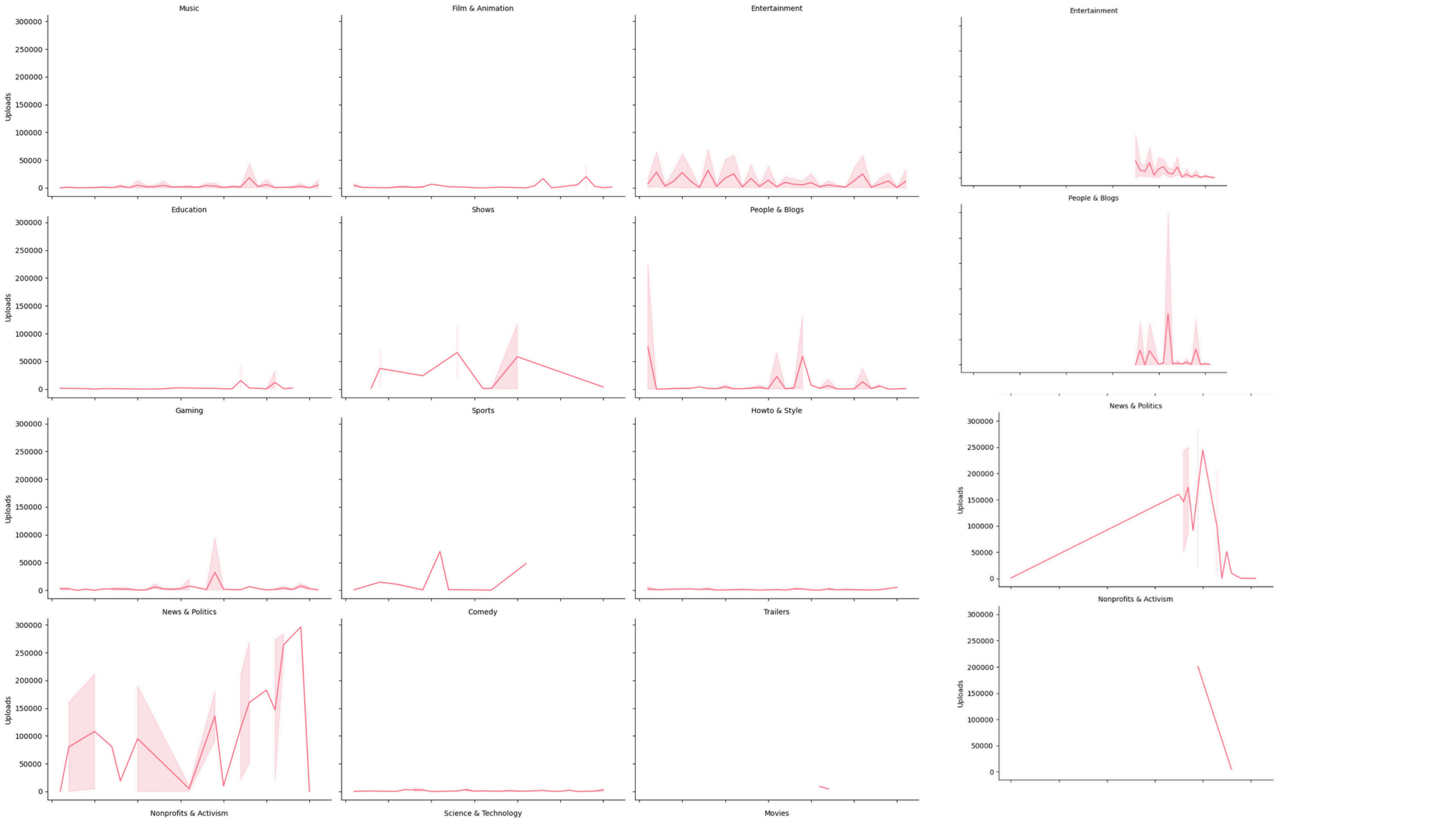
19. Are there any seasonal trends in the number of videos uploaded by YouTube channels?



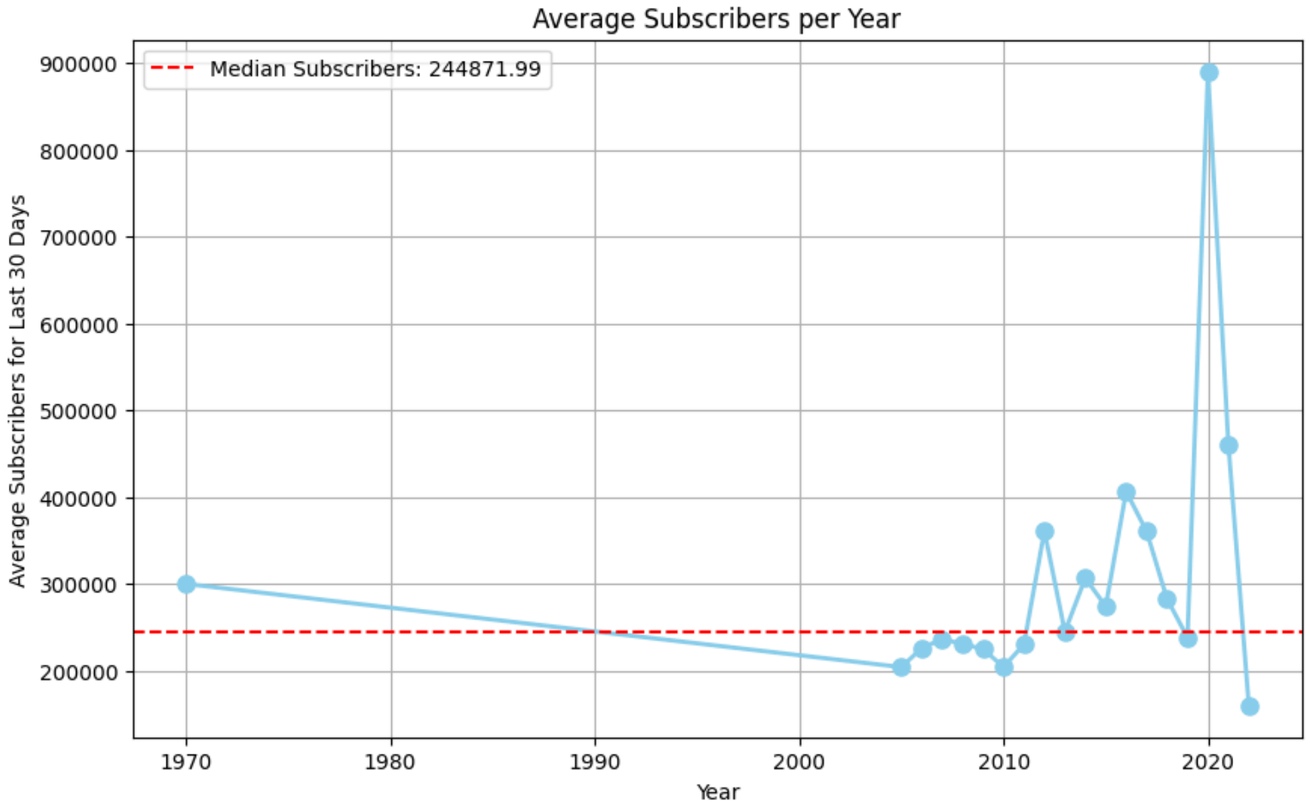



```
count        18.000000
mean      17500.490944
std       33571.467728
min         766.000000
25%        2167.030571
50%        3557.900000
75%        9767.997961
max      112484.384615
Name: uploads, dtype: float64
```

The availability of average upload data shows a significant lack from 1970 to 2005, likely due to limited digital content platforms at the time. Post-2005, there's a steady increase in uploads until around 2010, possibly reflecting broader internet adoption and content creation. A sharp decline follows, possibly due to shifts in platform policies or user behavior. Peaks in 2018 and subsequent dips suggest dynamic trends influenced by evolving technology, user preferences, and platform dynamics shaping digital content dissemination.


Music
Film & Animation
Entertainment
Entertainment
Education
Shows
People & Blogs
People & Blogs
Gaming
Sports
Howto & Style
News & Politics
News & Politics
Comedy
Trailers
Nonprofits & Activism
Nonprofits & Activism
Science & Technology
Movies
Uploads
0
50000
100000
150000
200000
250000
300000

20. What is the average number of subscribers gained per month since the creation of YouTube channels till now?



The observed trends in average subscribers per year indicate a lack of data between 1970 and 2005, likely reflecting the early stages of digital content creation. Subsequent increases around 2012 and 2020 suggest both biased dataset collection towards recent years and increased internet accessibility influencing higher subscriber growth rates, reflecting evolving digital engagement patterns over time.

THANKYOU