Media & Technology : YouTube Dataset

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

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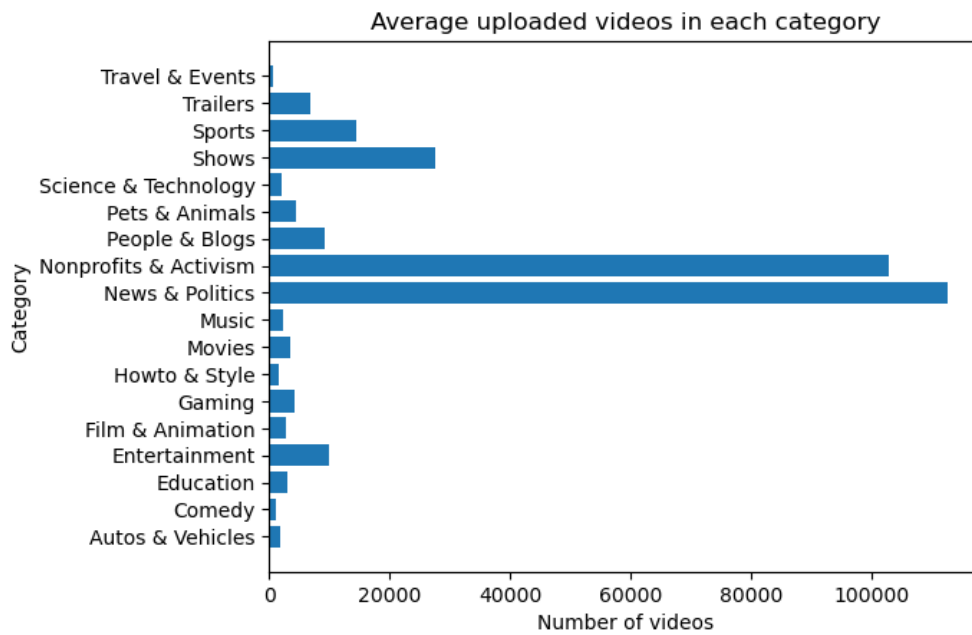
2. Which category has the highest average number of subscribers?

Ans: Category is 'Shows' and number of subscribers is 41615384.6

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

```
Category

Autos & Vehicles         1550.666667
Comedy                   1202.557143
Education                3087.086957
Entertainment            9969.650685
Film & Animation         2899.340909
Gaming                   4198.744681
Howto & Style            1695.500000
Movies                   3553.000000
Music                    2325.945813
News & Politics        112484.384615
Nonprofits & Activism  102912.000000
People & Blogs           9256.793893
Pets & Animals           3562.800000
Science & Technology     2114.058824
Shows                   27443.692308
Sports                  19129.833333
Trailers                 6839.000000
Travel & Events           766.000000
```
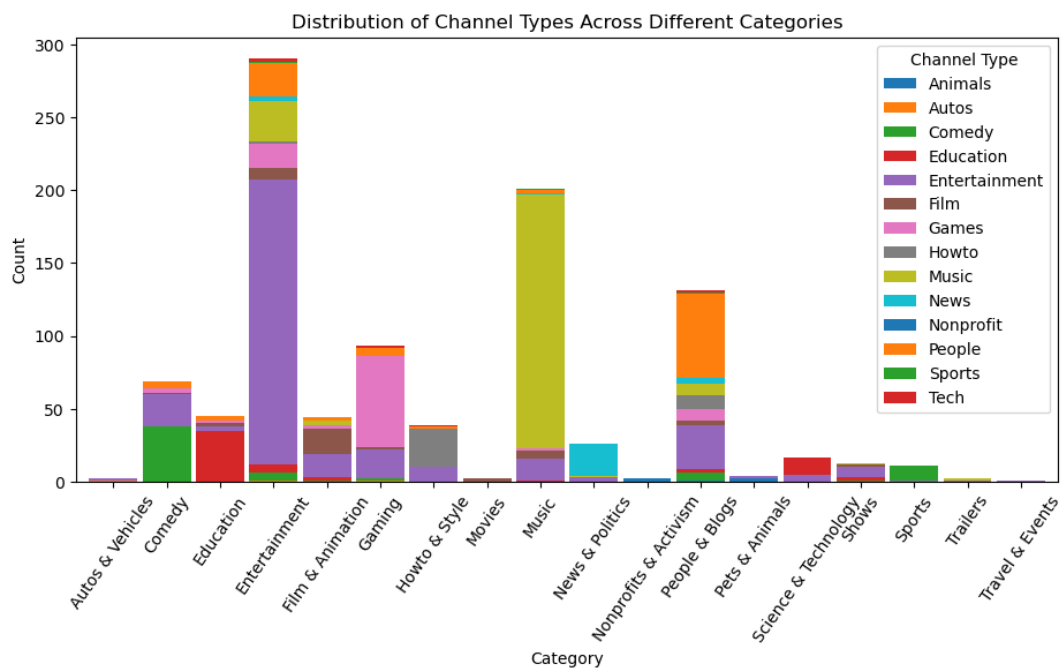
Average uploaded videos in each category

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

```
Country                 Number of Youtube channels
United States                   440
 India                          169
 Brazil                         62
 United Kingdom                 44
 Mexico                         33
```
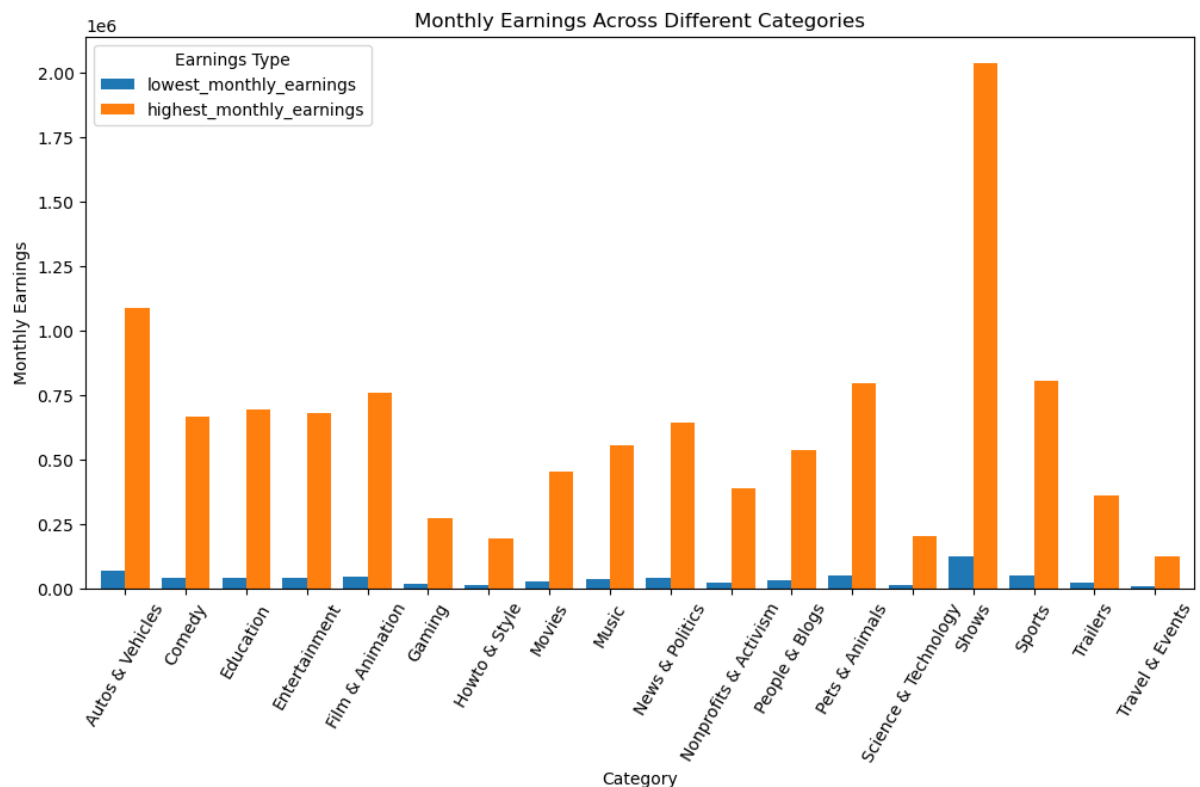
5. What is the distribution of channel types across different categories?

Distribution of Channel Types Across Different Categories

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

   Ans: Yes, as the correlation coefficient is 0.75, there is moderate-to-good correlation between no. of subscribers and video views.

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

8. **What is the overall trend in subscribers gained in the last 30 days across all channels?**
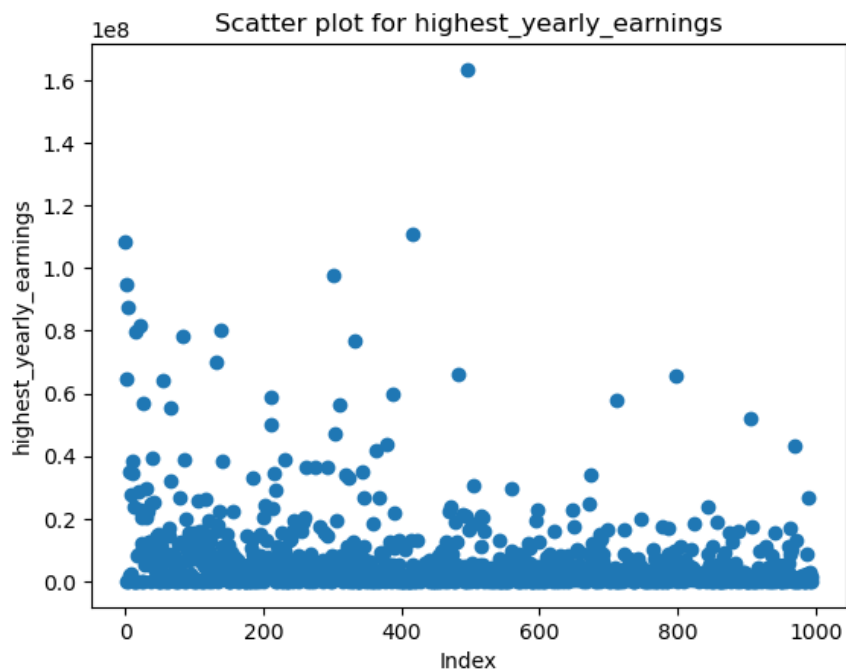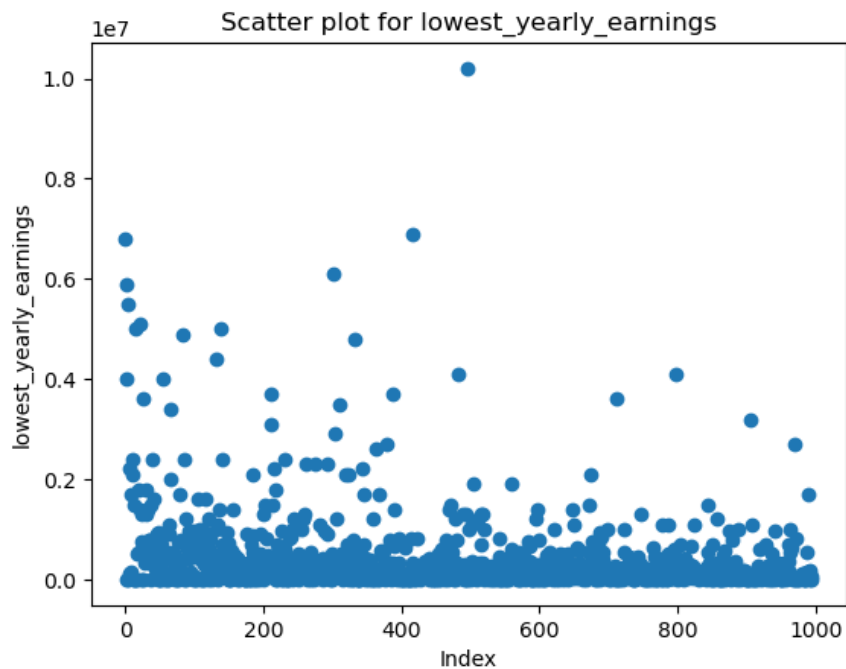
```
count    9.920000e+02
mean     2.990868e+05
std      5.051075e+05
min      1.000000e+00
25%      1.000000e+05
50%      2.000000e+05
75%      2.000000e+05
max      8.000000e+06
```

- On **average**, each channel gained approximately **299,086** subscribers in last 30 days
- Standard deviation value indicates **high variability** in the number of subscribers gained
- The **minimum** number of subscribers gained is **1** for ofcourse the channel with least growth
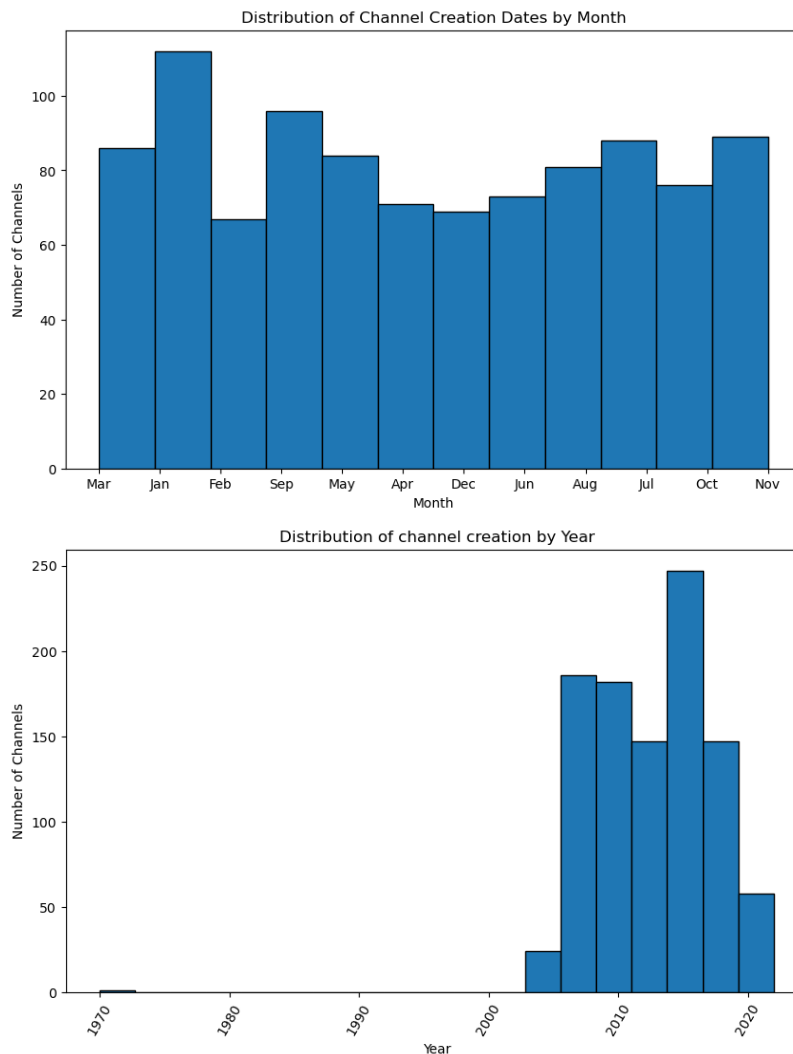
- The **maximum** number of subscribers gained is **8000,000** i.e., **8 million** for the channel with highest growth
- The first quartile i.e., **25%** of the channels gained **100,000** subscribers or lesser
- The **median** number of subscribers gained is **200,000** i.e., half of the channels gained **2 lakh** subscribers
- **75%** of the channels gained **2 lakh** subscribers or lesser

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

Scatter plot for lowest_yearly_earnings

Scatter plot for highest_yearly_earnings

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

Distribution of Channel Creation Dates by Month

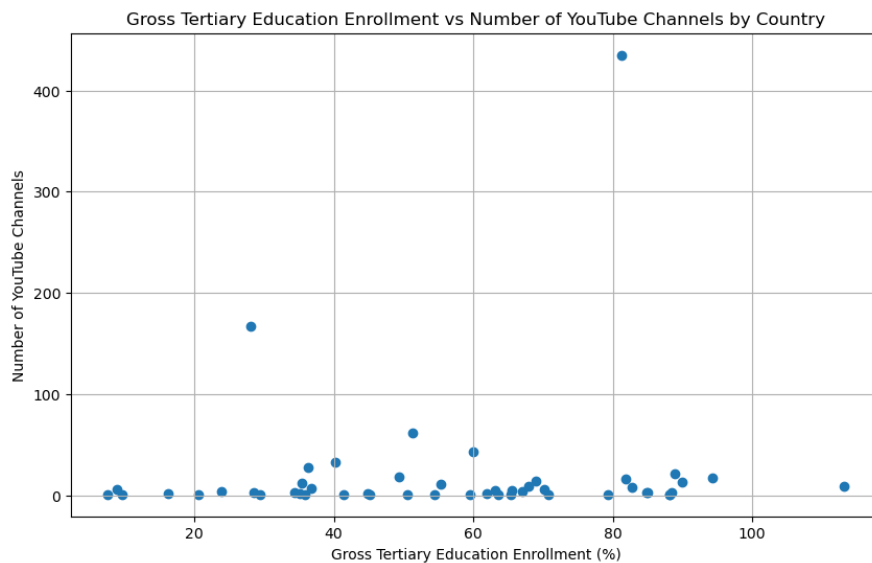Distribution of channel creation by Year

Ans: There is no clear or consistent trend in channel creation over days of the month. No. channels created vary considerably.

There is not a proper trend in channel creation over months, but comparably in Jan more no. of channels are created.
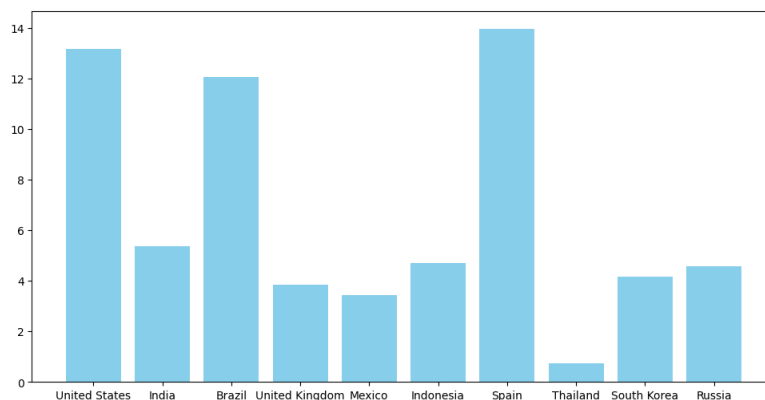
The earliest channel is ofcourse of youtube in 1970. After 2005, the creation of channels increased gradually. In 2006, it shoot up and then became somewhat stable. Around 2013-14 it again took a higher jump and then decreased slowly.

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

Gross Tertiary Education Enrollment vs Number of YouTube Channels by Country

Ans: The correlation coefficient (0.09) between the two is less, implying close to no relation between gross tertiary education enrolment and no. of youtube channels in a country. Also from the scatter plot you can see that the points are widely spread with no clear pattern. Significant outliers like countries with 400+ channels having gross tertiary education enrolment of around 80%

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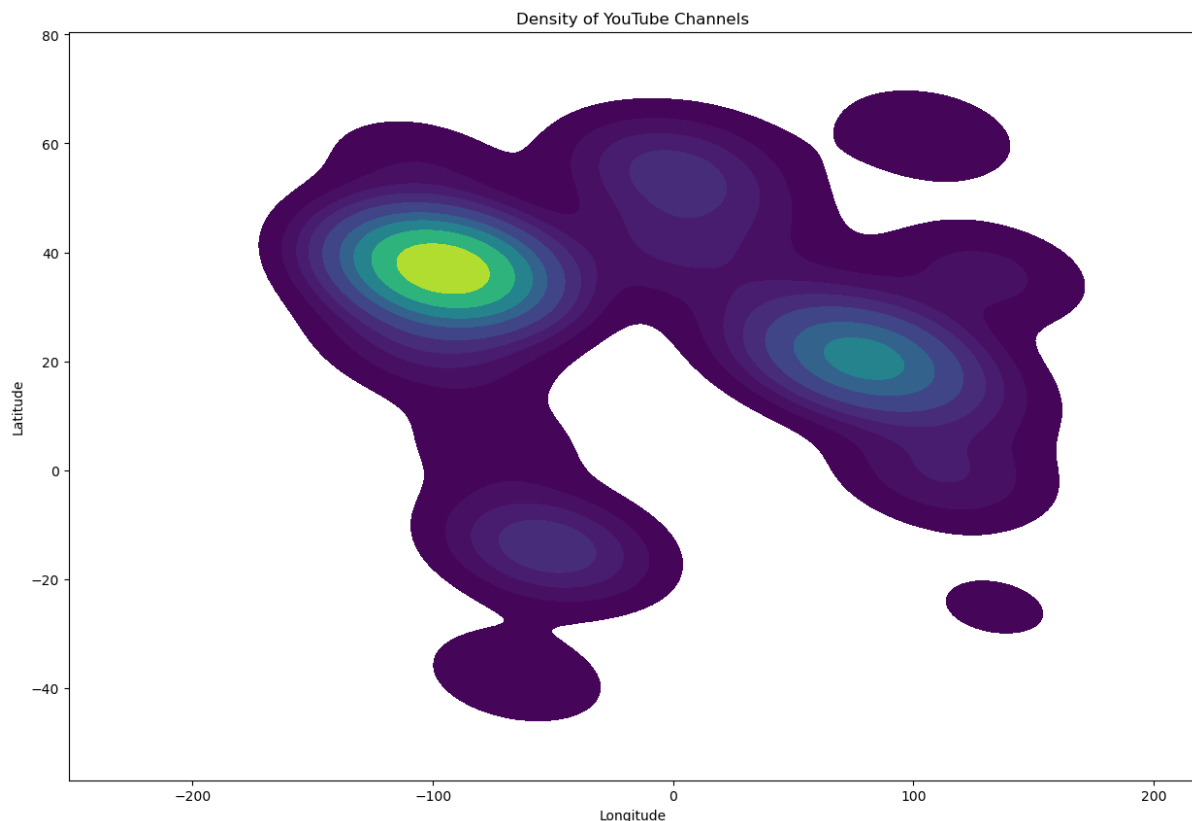

- There is not a proper trend or pattern in the unemployment rate for the top 10 countries with highest no. of channels.
- US being the country with highest no. of channels having the unemployment rate close to 13%.
- Spain being the country having highest unemployment rate close to 14%.

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

```
          Country  avg percentage
0     Afghanistan        0.270811
1         Andorra        7.481535
```

2	Argentina	1.142688
3	Australia	0.603822
4	Bangladesh	1.685784
5	Barbados	0.002472
6	Brazil	5.065076
7	Canada	0.846618
8	Chile	0.459129
9	China	23.299968
10	Colombia	1.128528
11	Cuba	0.241563
12	Ecuador	0.307283
13	Egypt	1.185705
14	El Salvador	0.129769
15	Finland	0.130382
16	France	1.496052
17	Germany	1.778036
18	India	13.020023
19	Indonesia	4.187958
20	Iraq	0.767974
21	Italy	1.178965
22	Japan	3.200401
23	Jordan	0.254663
24	Kuwait	0.116290
25	Latvia	0.036071
26	Malaysia	0.676547
27	Mexico	2.836761
28	Morocco	0.635064
29	Netherlands	0.440184
30	Pakistan	2.209324
31	Peru	0.701827
32	Philippines	1.409051
33	Russia	2.976545
34	Samoa	0.000984
35	Saudi Arabia	0.796292
36	Singapore	0.157655
37	South Korea	1.163893
38	Spain	1.048371
39	Sweden	0.249359
40	Switzerland	0.175038
41	Thailand	0.975596
42	Turkey	1.744119
43	Ukraine	0.852345
44	United Arab Emirates	0.234393
45	United Kingdom	1.545390
46	United States	7.481535
47	Venezuela	0.695526
48	Vietnam	0.976634

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

Density of YouTube Channels

- The part with yellow shaded region is US and is also correct because we know that highest number of channels are from the country US

- The greenish part region somewhere around latitude(20) and longitude(90) is Indian which makes sense as it is 2nd most country having a large no. of youtube channels after US
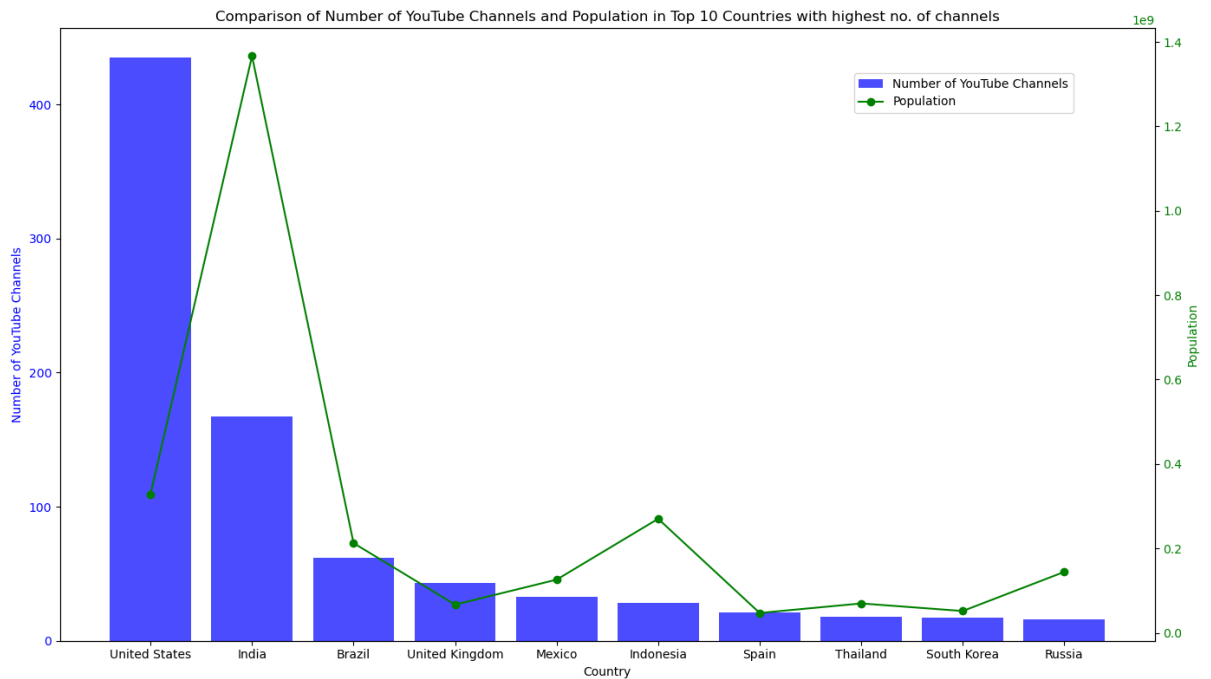
15. What is the correlation between the number of subscribers and the population of a country?

Ans: Correlation coefficient is 0.35436488224143614
Implies a moderate correlation between number of subscribers and population of a country.

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

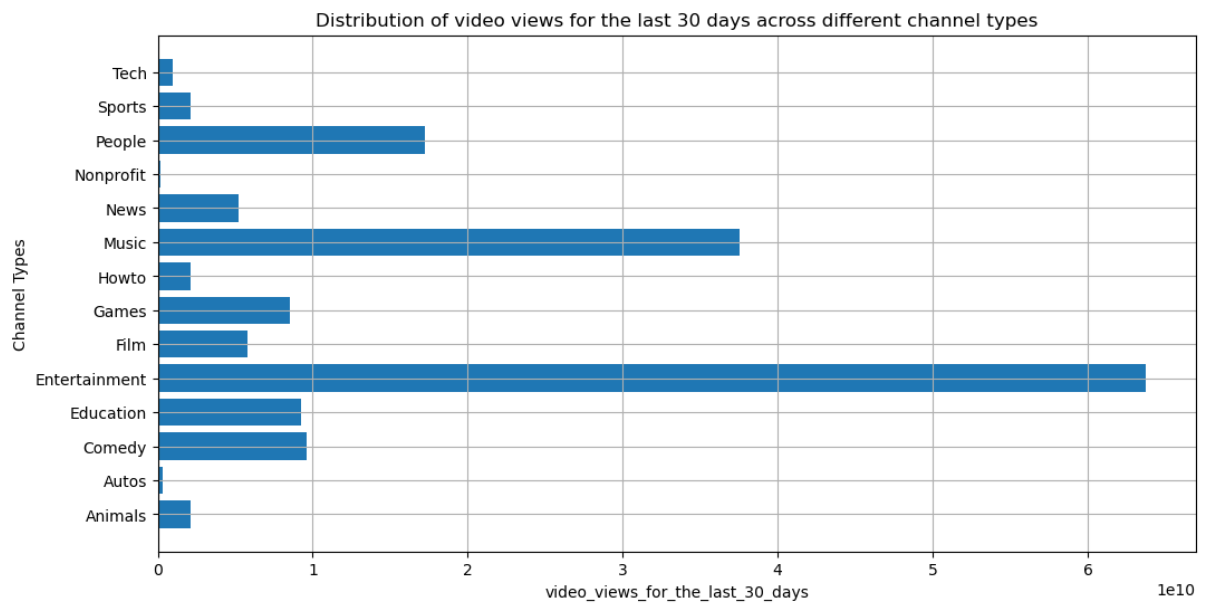Comparison of Number of YouTube Channels and Population in Top 10 Countries with highest no. of channels

- The United States has a disproportionately high number of YouTube channels relative to its population.
- India, with the highest population, has significantly fewer YouTube channels than the United States, highlighting a disparity in digital content creation.
- Countries like India, Indonesia, Brazil, and Mexico with large populations but relatively fewer YouTube channels, could be potential growth markets for YouTube content creation.
- Western countries (US, UK) have a higher concentration of YouTube channels compared to Eastern countries (India, Indonesia) when adjusted for population.

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

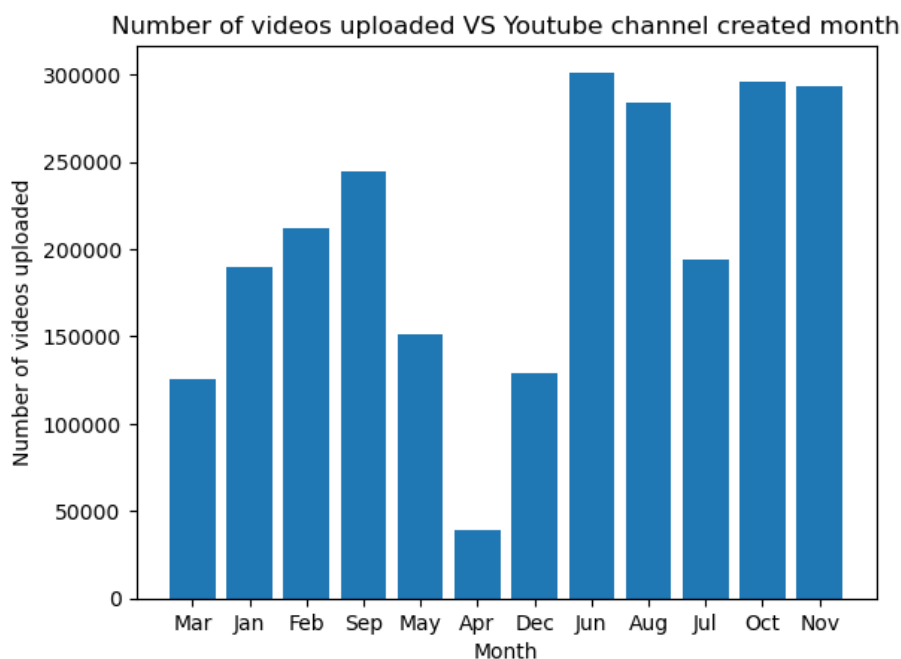   Ans: Correlation coefficient = 0.21139636814187898
    Implies a weak correlation between number of subscribers gained in the last 30 days and unemployment rate in a country.

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

Distribution of video views for the last 30 days across different channel types

- **Entertainment** and **Music** dominate in terms of views, which aligns with the general trend of these types of content being highly engaging and widely shared.
- Categories like **Tech, Autos,** and **Nonprofit** have smaller view counts but cater to more specific niche audiences who are interested in these topics.
- The diversity in view counts across different categories indicates a broad spectrum of interests among YouTube viewers.

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

Number of videos uploaded VS Youtube channel created month

- High Activity Periods: Summer and Fall (June, August, October, November): These months generally show higher activity in terms of video uploads. This might be due to the fact that summer months often see more leisure time and vacations, leading to

more content creation. Fall months might see a spike due to the return to routine schedules, potentially more planning and execution of content.

- Low Activity Periods: December and May: These months are outliers with significantly lower activity. December might see a dip due to the holiday season, where people are less focused on content creation.
  May could be a transitional month, marking the end of spring and the beginning of summer, leading to less activity.
- Transitions: April: Being just before summer, April might be a planning phase for many creators, leading to fewer uploads.
  July: Despite being a summer month, it shows lower uploads compared to June and August, which could indicate mid-summer vacations or a temporary lull in activity.

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

Ans : 184875

---------- END -----------