Media & Technology

CASE STUDY

PRAKRITI PANDEY

PREPROCESSING THE DATA

- Read the dataset
- Handle missing values
  - Replaces string versions of missing values ('nan', '', 'NAN', 'NaN') with proper np.nan (so pandas recognizes them as missing).
- Clean column names
  - .str.strip(): Removes extra spaces at the beginning and end of column names.
  - .str.lower(): Converts all column names to lowercase.
  - .str.replace(" ", "_"): Replaces spaces with underscores for consistency
- Ensure numeric columns are actually numeric
  - Defines a list numeric_cols containing column names expected to hold numeric values (like subscribers, views, population, unemployment rate, etc.).
  - For each of these columns (if it exists in the dataset):
    - Converts the column to numeric values using pd.to_numeric().
    - If conversion fails (e.g., column has text instead of numbers), it replaces the invalid entries with NaN (errors='coerce').

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

Output:

```
                    youtuber   subscribers
0                    T-Series   245000000.0
1               YouTube Movies   170000000.0
2                     MrBeast   166000000.0
3   Cocomelon - Nursery Rhymes   162000000.0
4                    SET India   159000000.0
5                       Music   119000000.0
6          ýýý Kids Diana Show   112000000.0
7                    PewDiePie   111000000.0
8                  Like Nastya   106000000.0
9                Vlad and Niki    98900000.0
```

Takeaways:

- 'T-Series' and 'Mr Beast' dominate the list with the highest number of subscribers.

- Children's content channels (Cocomelon, Kids Diana Show, Like Nastya, Vlad and Niki) dominate the list.

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

Output:

```
category
Shows    4.161538e+07
Name: subscribers, dtype: float64
```
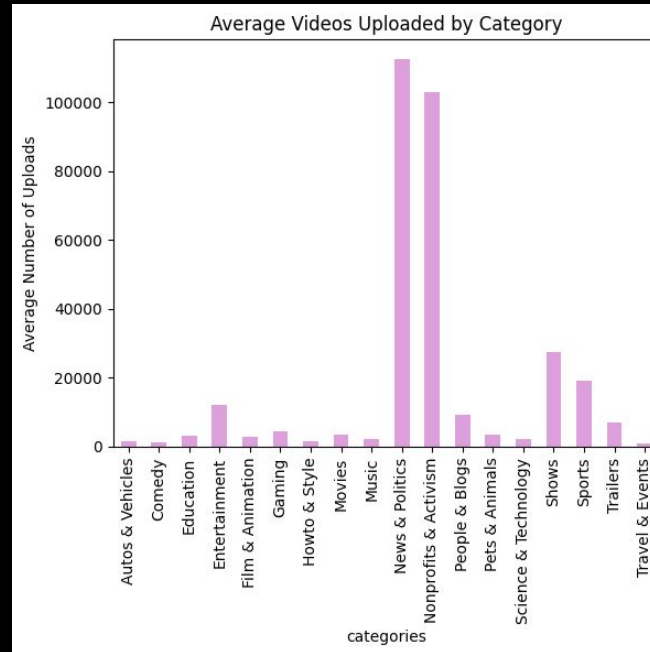
Explanation:

- Among all YouTube categories, "Shows" has the highest average number of subscribers per channel.
- On average, a channel in the Shows category has around 41.6 million subscribers.

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

Output:

Autos & Vehicles	1550.666667
Comedy	1202.557143
Education	3087.086957
Entertainment	12052.445378
Film & Animation	2861.844444
Gaming	4285.273684
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 Videos Uploaded by Category

Takeaways:

● News & Politics and Nonprofits & Activism have extremely high averages
● Shows and Sports also have very high average uploads.
● Categories like Travel & Events and Comedy have relatively fewer uploads.

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

Output:

```
country
United States       315
India               169
Brazil               62
United Kingdom       44
Mexico               33
Name: youtuber, dtype: int64
```

Takeaways:

- The United States has the most YouTube channels in the dataset (315).

- India is 2nd with 169 channels.

- Brazil, UK, and Mexico follow, but with significantly fewer channels.

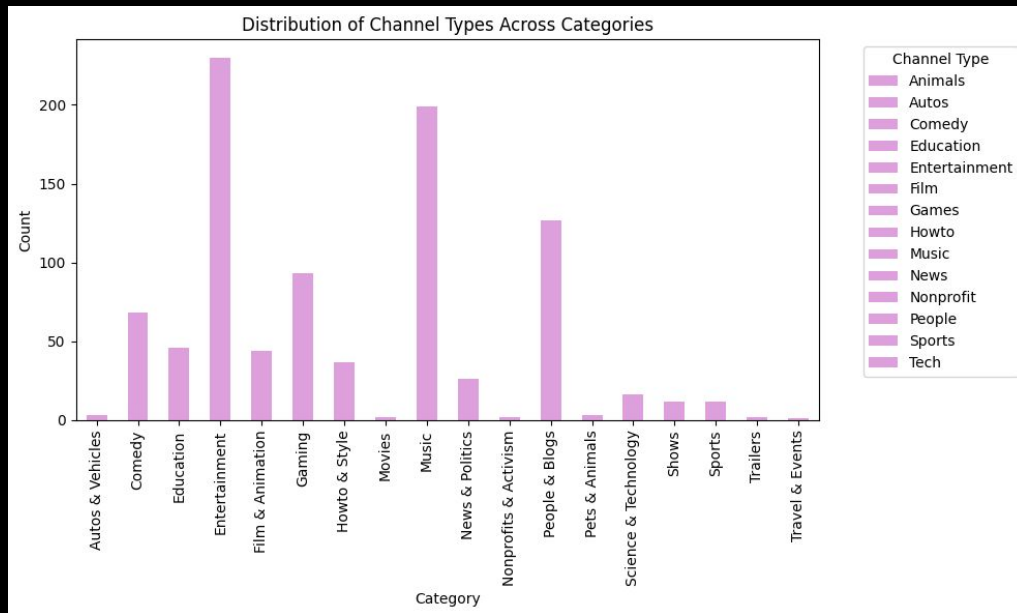- Overall, USA and India dominate the YouTube creator chart.

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

Output:

```
Distribution of channel types across different categories

 channel_type       Animals  Autos  Comedy  Education  Entertainment  Film
category
Autos & Vehicles         0      2       0          0              1     0
Comedy                   0      0      39          0             20     1
Education                0      0       0         36              3     2
Entertainment            0      1       5          2            168     6
Film & Animation         0      0       1          2             16    17
Gaming                   0      1       1          0             18     2
Howto & Style            0      0       0          0              7     0
Movies                   0      0       0          0              0     2
Music                    0      0       0          1             12     5
News & Politics          0      0       0          0              3     0
Nonprofits & Activism    0      0       0          0              0     0
People & Blogs           1      0       5          3             26     3
Pets & Animals           2      0       0          0              1     0
Science & Technology     0      0       0          0              4     0
Shows                    0      0       1          2              6     2
Sports                   0      0       0          0              1     0
Trailers                 0      0       0          0              1     0
Travel & Events          0      0       0          0              1     0
```


Distribution of Channel Types Across Categories

channel_type	Games	Howto	Music	News	Nonprofit	People	Sports
category							
Autos & Vehicles	0	0	0	0	0	0	0
Comedy	3	0	0	0	0	5	0
Education	2	0	0	0	0	3	0
Entertainment	11	0	22	3	0	11	0
Film & Animation	3	0	3	0	0	2	0
Gaming	64	0	0	0	0	6	0
Howto & Style	0	27	0	0	0	2	0
Movies	0	0	0	0	0	0	0
Music	2	0	175	1	0	2	1
News & Politics	0	0	1	22	0	0	0
Nonprofits & Activism	0	0	0	0	2	0	0
People & Blogs	8	9	8	4	0	58	1
Pets & Animals	0	0	0	0	0	0	0
Science & Technology	0	0	0	0	0	0	0
Shows	0	0	1	0	0	0	0
Sports	0	0	0	0	0	0	11
Trailers	0	0	1	0	0	0	0
Travel & Events	0	0	0	0	0	0	0

channel_type	Tech
category	
Autos & Vehicles	0
Comedy	0
Education	0
Entertainment	1
Film & Animation	0
Gaming	1
Howto & Style	1
Movies	0
Music	0
News & Politics	0
Nonprofits & Activism	0
People & Blogs	1
Pets & Animals	0
Science & Technology	12
Shows	0
Sports	0
Trailers	0
Travel & Events	0

Takeaways:

- Entertainment category is the most diverse: it has channels across many types (Entertainment, Music, Games, etc.).
- Gaming category is dominated by Games-type channels (64).
- Music category is dominated by Music-type channels (175).
- Science & Technology category is small but mostly Tech-type channels (12).
- Some categories like Nonprofits & Activism have very few or no channels in most types.

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

Output:

```
Entertainment            238
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
Movies                     2
Nonprofits & Activism      2
Trailers                   2
Travel & Events            1
Name: youtuber, dtype: int64
correlation between the number of subscribers and total video
views for YouTube channels:  0.7481786016237689
```
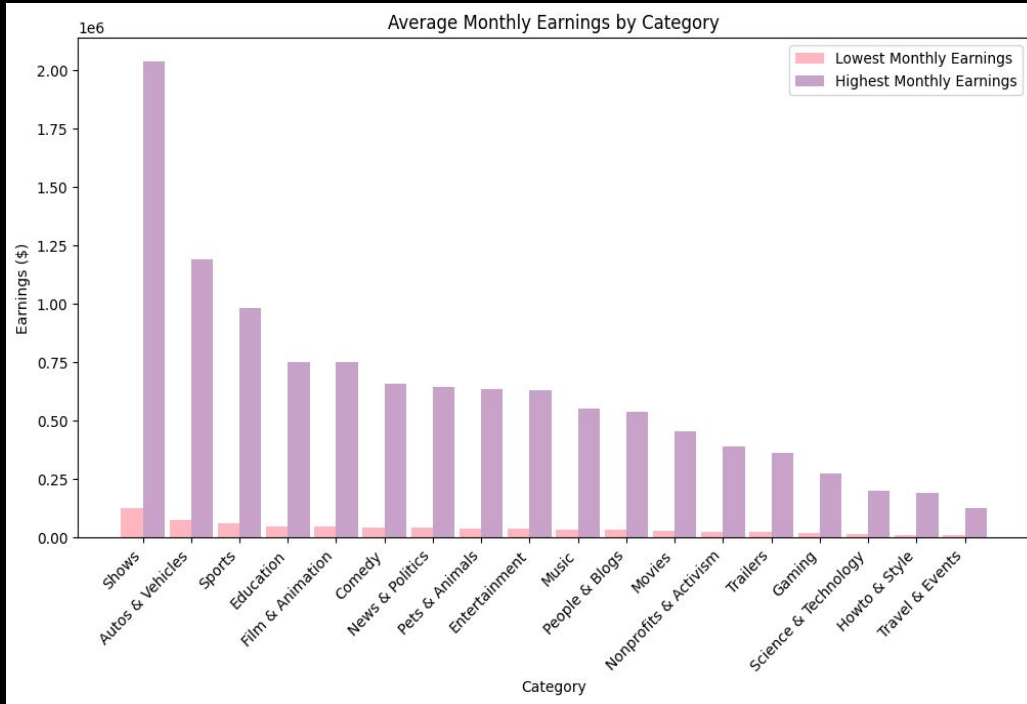
Takeaways:

- **The correlation is ~0.75, which is a strong positive correlation.**

- This means:
  - Channels with more subscribers usually also have more total video views.
  - But it's not a perfect 1.0, so there are exceptions (e.g., a channel could have fewer subscribers but very viral videos).

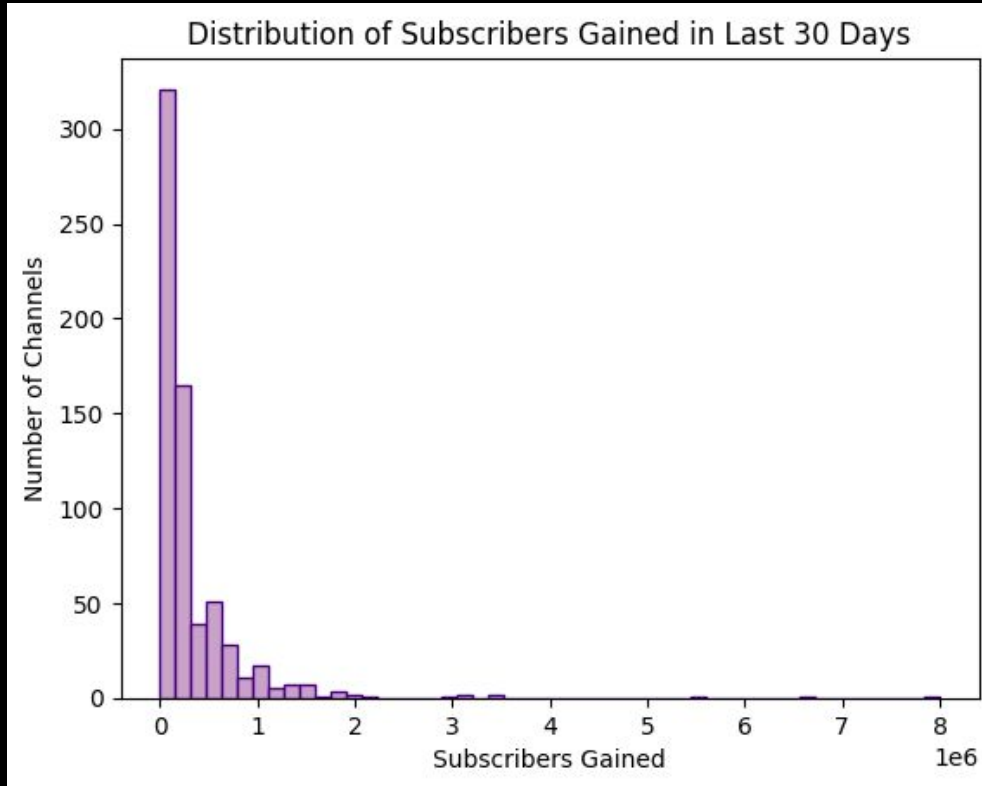# 7. How do the monthly earnings vary throughout different categories?

Average Monthly Earnings by Category

Takeaways:

- Shows lead by far in earnings, with Autos & Vehicles and Sports also very high.

- Education, Entertainment, Comedy, Film & Animation, News & Politics, and Pets & Animals fall in the mid-range. Music,

- People & Blogs, Movies, and Nonprofits & Activism earn moderately.

- Gaming, Science & Tech, Howto & Style, and Travel & Events are lowest, showing an uneven income distribution where top categories dominate.

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

Distribution of Subscribers Gained in Last 30 Days

Takeaways:

● Most channels gained relatively few subscribers in the last 30 days (close to 0–200k).

● A small number of channels gained very high subscribers (in the millions).

● This shows the distribution is highly skewed therefore subscriber growth is uneven

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

Output:

```
Outliers with unusually high earnings:

                         youtuber  highest_yearly_earnings
0                         T-Series              108400000.0
2                          MrBeast               64700000.0
3        Cocomelon - Nursery Rhymes             94800000.0
4                         SET India              87500000.0
6              ýýý Kids Diana Show               35100000.0
..                             ...                      ...
908                    Susy Mouriz               17700000.0
970                       _vector_               43400000.0
990                  Natan por Aï¿                26500000.0
995                            WWE               34300000.0
999            Infobells - Hindi                 39200000.0

[101 rows x 2 columns]
Outliers with unusually low earnings:
 Empty DataFrame
Columns: [youtuber, highest_yearly_earnings]
Index: []
```
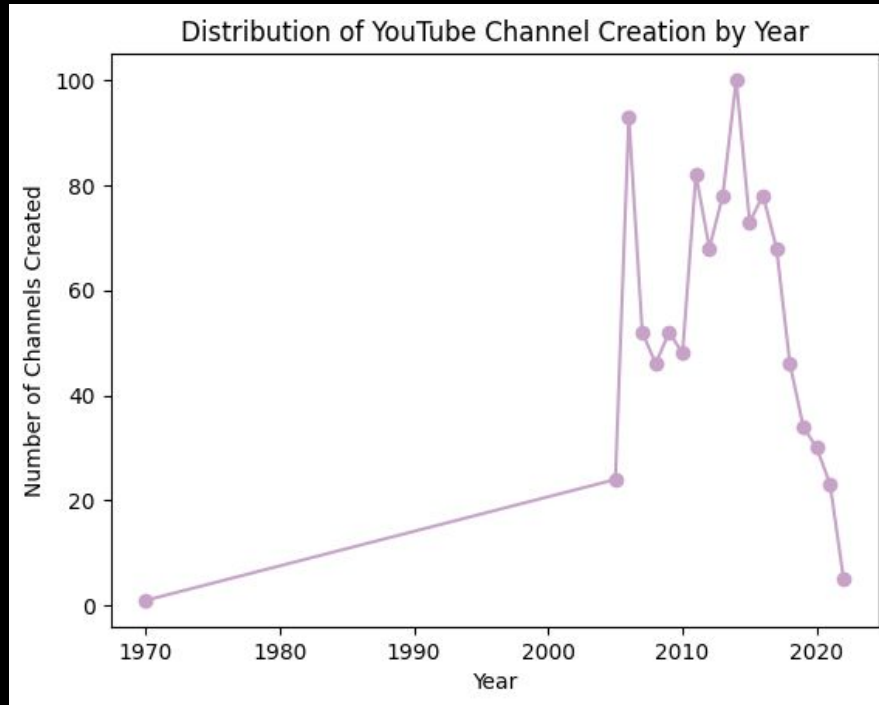
Takeaways:

- There are only high-earning outliers, no low-earning ones.
- A few top channels (like T-Series, Cocomelon, SET India, MrBeast) make tens of millions more per year than the average channel.
- Most other channels earn far less, but they're still within a normal range — not extreme outliers.

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

Output:

```
Channels created by year:
 created_year
1970.0      1
2005.0     24
2006.0     93
2007.0     52
2008.0     46
2009.0     52
2010.0     48
2011.0     82
2012.0     68
2013.0     78
2014.0    100
2015.0     73
2016.0     78
2017.0     68
2018.0     46
2019.0     34
2020.0     30
2021.0     23
2022.0      5
Name: count, dtype: int64
```


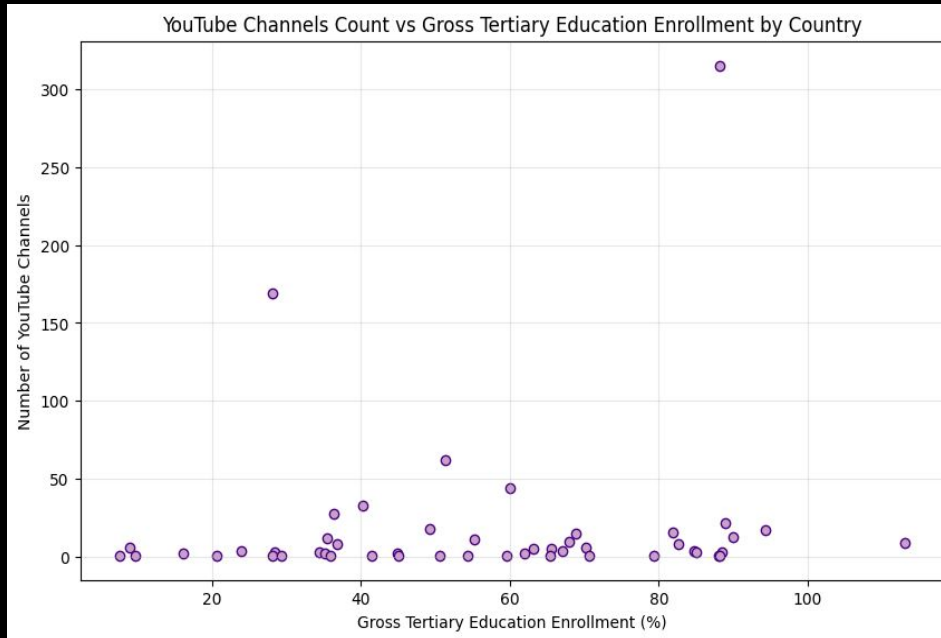Distribution of YouTube Channel Creation by Year

Takeaways:

- Only one of the channels were created between 1970 - 2004
- General rise in channel creation from 2005 to 2014,
- Peak around 2014
- Gradual decline after 2015. This may suggest there has been a recent slowdown in the youtube channel creation.

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

Output:

YouTube Channels Count vs Gross Tertiary Education Enrollment by Country

Takeaways:

- No clear linear relationship: Higher education enrollment does not consistently mean more YouTube channels.
- Instead, the data shows: Most countries have low channel counts, regardless of enrollment.
- Only a few countries stand out with unusually high numbers

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

Output:

```
Top 10 countries with
highest number of YouTube
channels:

United States        315
India                169
Brazil                62
United Kingdom        44
Mexico                33
Indonesia             28
Spain                 22
Thailand              18
South Korea           17
Russia                16
Name: count, dtype: int64
```
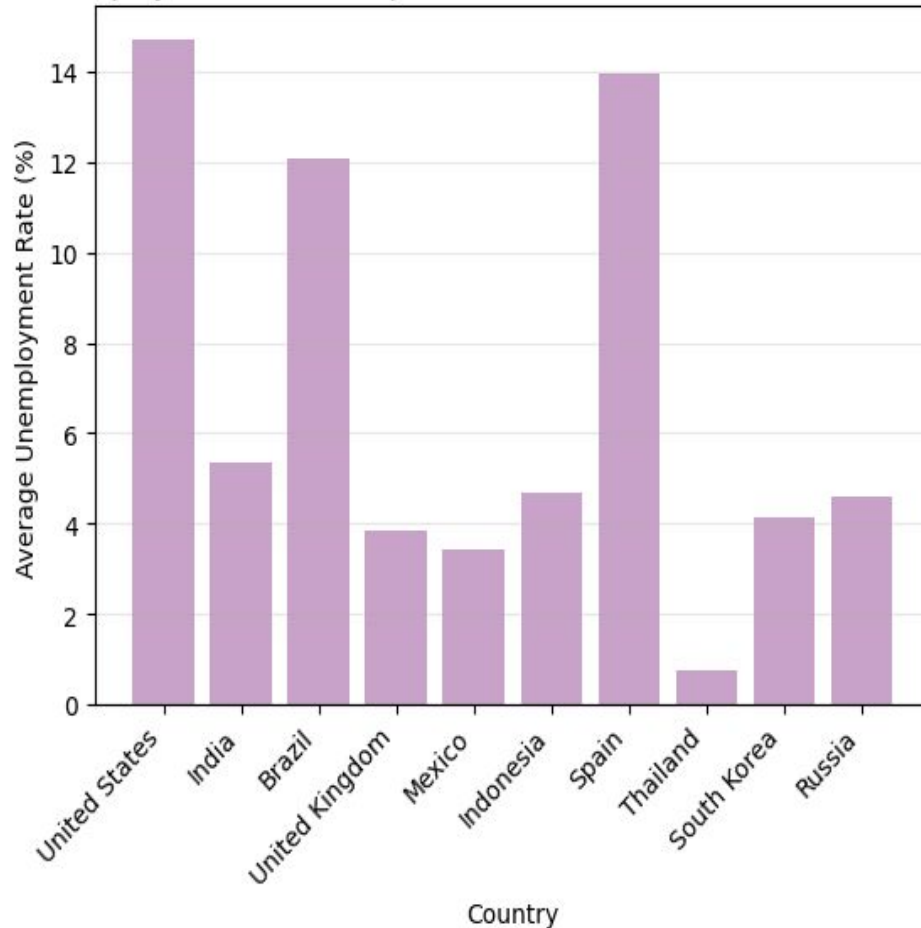
```
Unemployment rate for top 10
countries:
United States        14.70
India                 5.36
Brazil               12.08
United Kingdom        3.85
Mexico                3.42
Indonesia             4.69
Spain                13.96
Thailand              0.75
South Korea           4.15
Russia                4.59
Name: unemployment_rate,
dtype: float64
```

Unemployment Rate in Top 10 Countries with Most YouTube Channels

Takeaways:

- High unemployment countries (USA, Brazil, Spain) still have many channels

- Low unemployment countries (Thailand, UK, Mexico, South Korea, Russia) also appear in the top 10

- No clear trend: Both high and low unemployment nations dominate YouTube.

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

Output:

```
Average urban population percentage
across all countries with
YouTube channels:
223974718.82045454

Average urban population percentage
per country with YouTube channels:
Afghanistan              9797273.0
Andorra                        NaN
Argentina               41339571.0
Australia               21844756.0
Bangladesh              60987417.0
Barbados                   89431.0
Brazil                 183241641.0
Canada                  30628482.0
Chile                   16610135.0
China                  842933962.0
Colombia                40827302.0
Cuba                     8739135.0
Ecuador                 11116711.0
Egypt                   42895824.0
El Salvador              4694702.0
Finland                  4716888.0
France                  54123364.0
Germany                 64324835.0
India                  471031528.0
Indonesia              151509724.0
Iraq                    27783368.0
Italy                   42651966.0
Japan                  115782416.0
Jordan                   9213048.0
Kuwait                   4207083.0
Latvia                   1304943.0
Malaysia                24475766.0
Mexico                 102626859.0
Morocco                 22975026.0
Netherlands             15924729.0
```

```
Pakistan                79927762.0
Peru                    25390339.0
Philippines             50975903.0
Russia                 107683889.0
Samoa                      35588.0
Saudi Arabia            28807838.0
Singapore                5703569.0
South Korea             42106719.0
Spain                   37927409.0
Sweden                   9021165.0
Switzerland              6332428.0
Thailand                35294600.0
Turkey                  63097818.0
Ukraine                 30835699.0
United Arab Emirates     8479744.0
United Kingdom          55908316.0
United States          270663028.0
Venezuela               25162368.0
Vietnam                 35332140.0
india                  471031528.0
Name: urban_population, dtype: float64
```
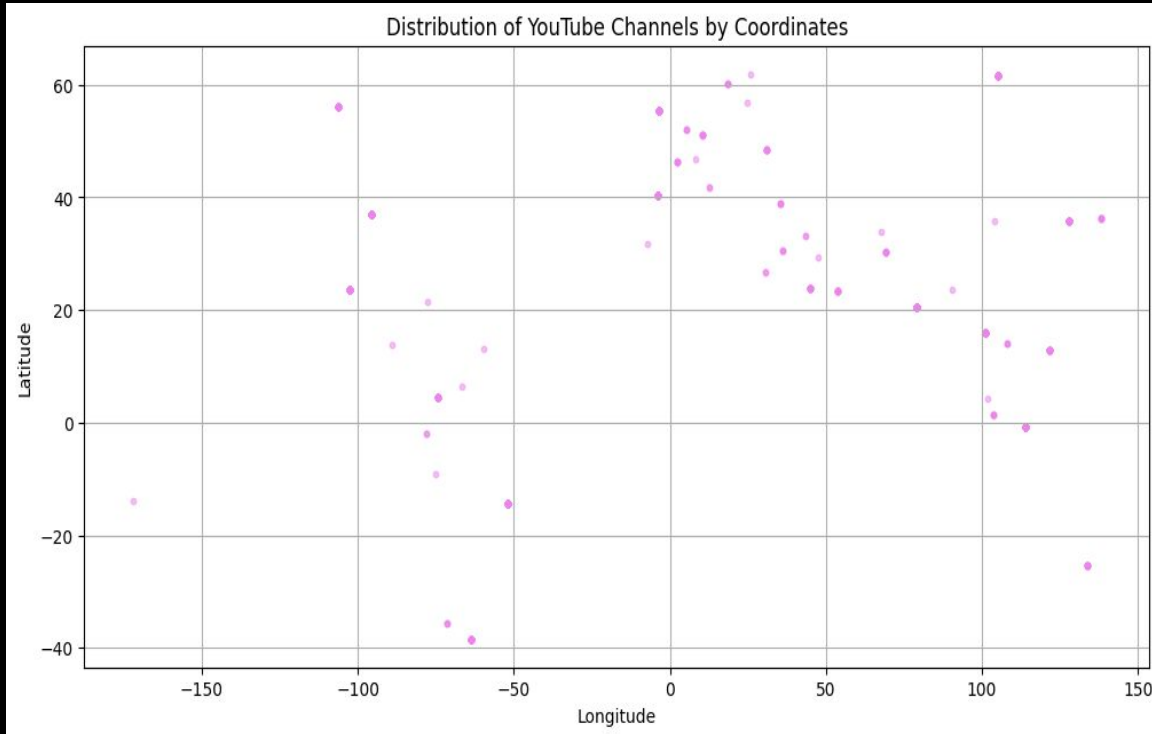
Takeaways:

- On average, countries with YouTube channels have a lot of people living in urban areas.

- This means that most YouTube activity is linked to countries with huge urban populations.

- The biggest contributors are countries like China, India, and the United States

- YouTube presence is strongly associated with highly urbanized and heavily populated countries.

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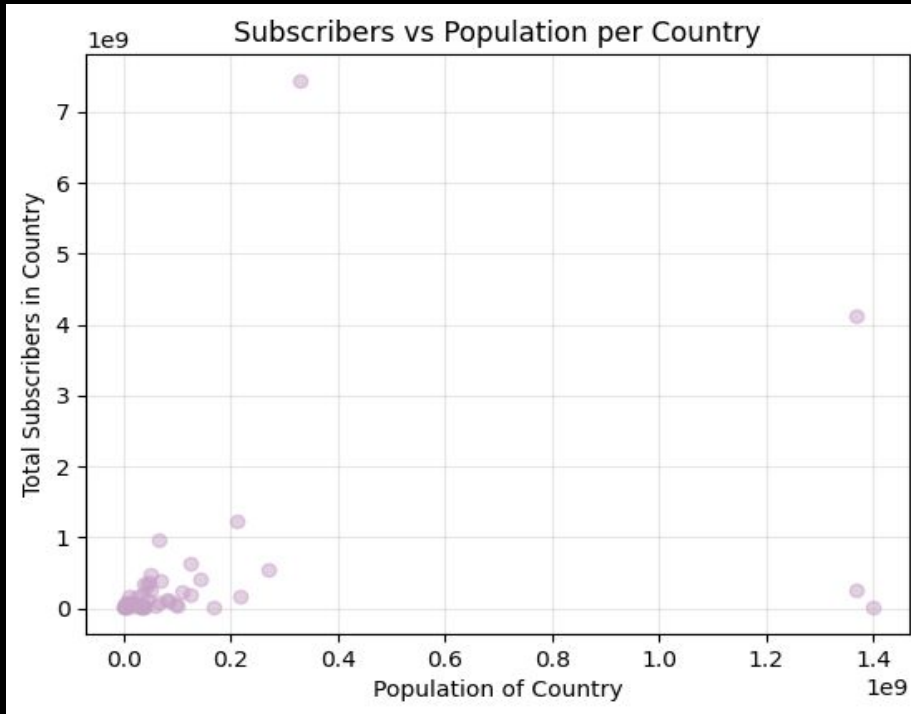

Distribution of YouTube Channels by Coordinates

Takeaways:
- Most channels are in the Northern Hemisphere, especially between latitudes 20°N to 60°N.

- Fewer dots are in the Southern Hemisphere (South America, parts of Africa, Oceania).

- Virtually no points in extreme latitudes (e.g., Arctic or Antarctic regions)

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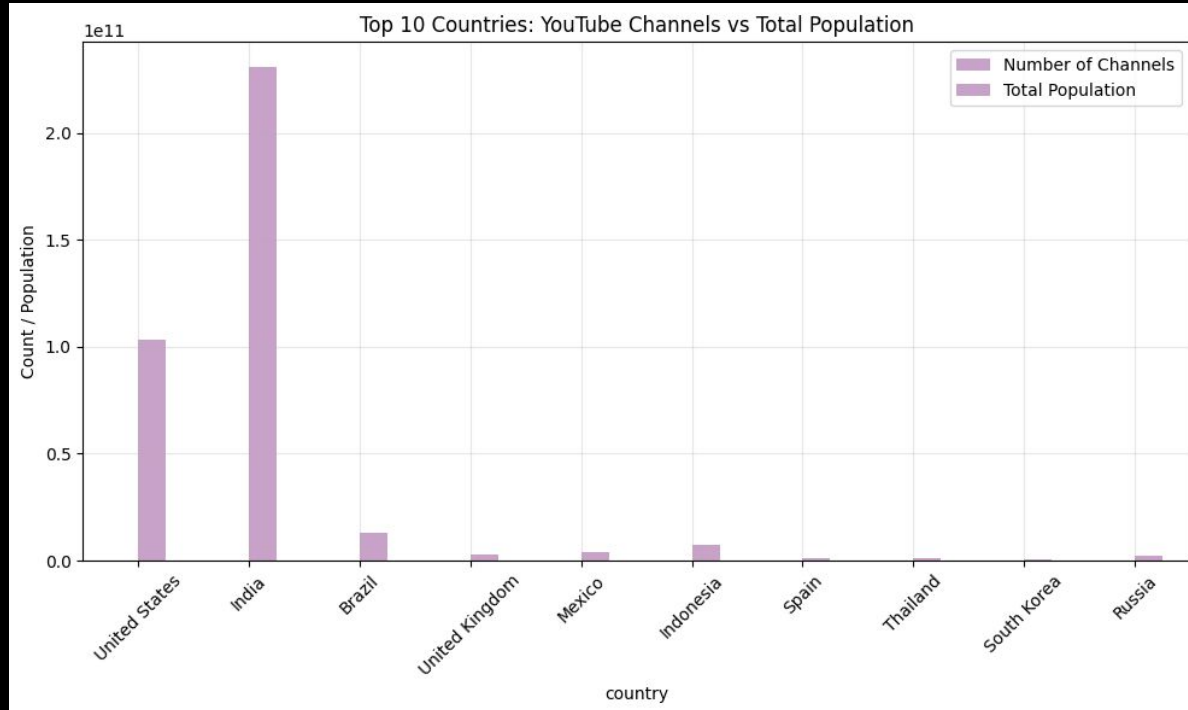
Output:

Subscribers vs Population per Country

Takeaways:

- There's no strong or clear positive correlation between population and the number of subscribers.
- Outliers are present. For example, 0.3B population but over 7B subscribers,
- For many smaller-population countries, subscriber counts cluster low
- This shows that number of subscribers isn't directly proportional to population

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

Output:

Top 10 Countries: YouTube Channels vs Total Population

United States	315
1.033954e+11	
India	169
2.309246e+11	
Brazil	62
1.317868e+10	
United Kingdom	44
2.940714e+09	
Mexico	33
4.158463e+09	
Indonesia	28
7.565710e+09	
Spain	22
Thailand	18
1.253260e+09	
South Korea	
8.790547e+08	
Russia	16
2.309977e+09	

Takeaways:

- Large countries (India, US) dominate absolute numbers.

- Some large-population countries (Indonesia, Mexico) lag behind in YouTube channel creation.

- This suggests that more opportunities of channel creation may come with a higher population.
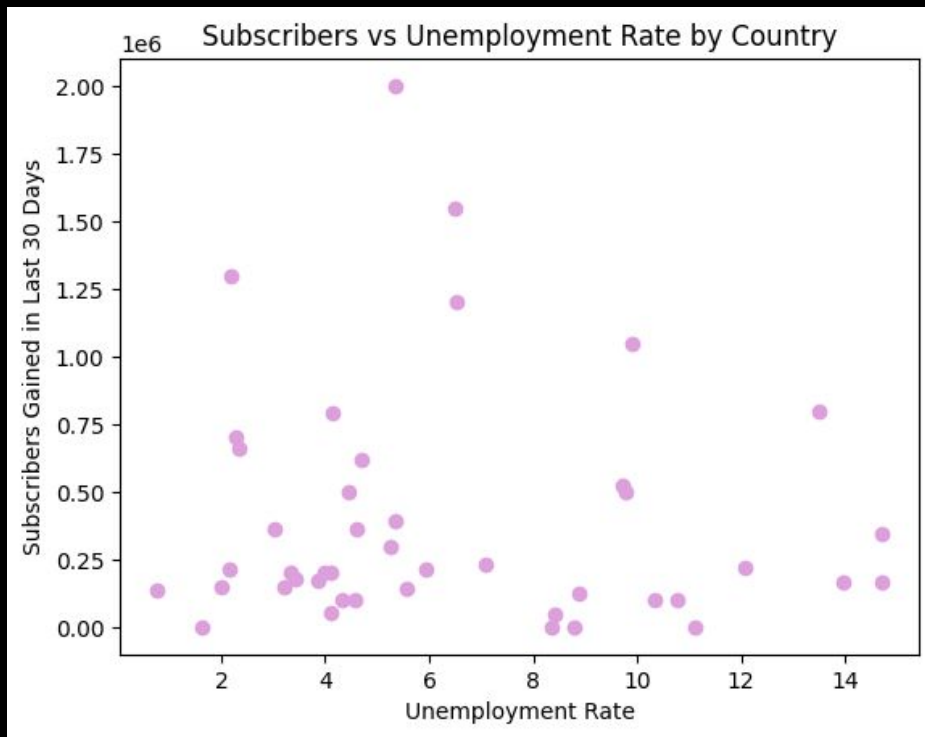
17. Is there a correlation between the number of subscribers gained in the last 30 days and the unemployment rate in a country?

Output:

```
number of subscribers gained in the last 30 days and the unemployment
rate in a country:                      subscribers_for_last_30_days
unemployment_rate
country
india                                   2.000000e+06                5.36
Sweden                                  1.550000e+06                6.48
Kuwait                                  1.300000e+06                2.18
Latvia                                  1.200000e+06                6.52
Italy                                   1.050000e+06                9.89
Turkey                                  8.000000e+05               13.49
South Korea                             7.888889e+05                4.15
Japan                                   7.000022e+05                2.29
United Arab Emirates                    6.625000e+05                2.35
Indonesia                               6.200001e+05                4.69
Colombia                                5.222222e+05                9.71
Pakistan                                5.000067e+05                4.45
Argentina                               5.000000e+05                9.79
India                                   3.898611e+05                5.36
Russia                                  3.642864e+05                4.59
Germany                                 3.607170e+05                3.04
United States                           3.446297e+05               14.70
Australia                               3.000005e+05                5.27
Chile                                   2.333347e+05                7.09
Brazil                                  2.181818e+05               12.08
Saudi Arabia                            2.145714e+05                5.93
```

```
Philippines                             2.125000e+05                2.15
El Salvador                             2.000000e+05                4.11
Ecuador                                 2.000000e+05                3.97
Malaysia                                2.000000e+05                3.32
Mexico                                  1.764706e+05                3.42
United Kingdom                          1.750404e+05                3.85
Jordan                                  1.666667e+05               14.72
Spain                                   1.636910e+05               13.96
Netherlands                             1.500000e+05                3.20
Vietnam                                 1.500000e+05                2.01
Canada                                  1.428573e+05                5.56
Thailand                                1.384615e+05                0.75
Ukraine                                 1.260020e+05                8.88
Egypt                                   1.000000e+05               10.76
Barbados                                1.000000e+05               10.33
Switzerland                             1.000000e+05                4.58
China                                   1.000000e+05                4.32
Singapore                               5.040025e+04                4.11
France                                  5.000350e+04                8.43
Venezuela                               1.000000e+02                8.80
Afghanistan                             1.200000e+01               11.12
Cuba                                    1.000000e+01                1.64
Samoa                                   3.000000e+00                8.36
Iraq                                             NaN               12.82
Morocco                                          NaN                9.02
Finland                                          NaN                6.59
Bangladesh                                       NaN                4.19
Peru                                             NaN                3.31
Andorra                                          NaN                 NaN
```

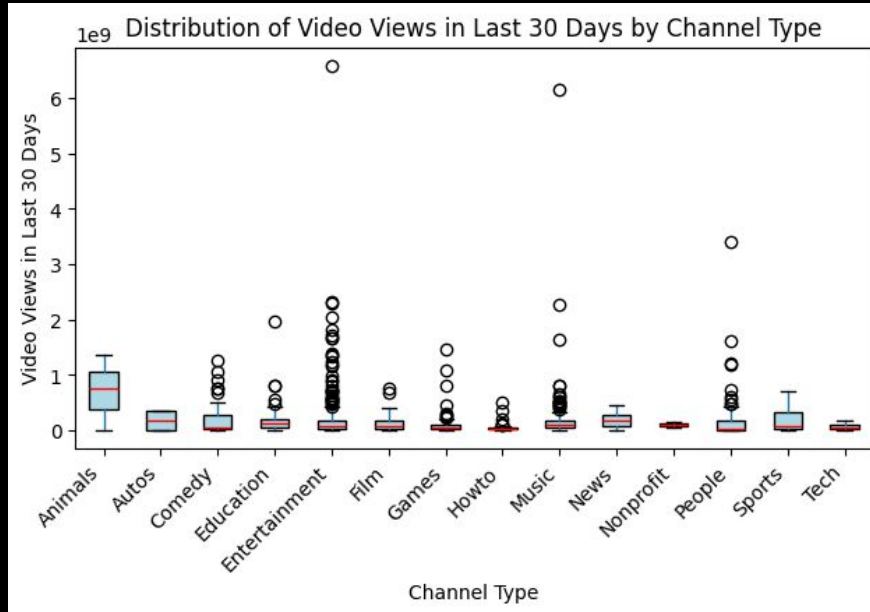Subscribers vs Unemployment Rate by Country

Takeaways:

- No clear correlation between unemployment and subscriber growth
- Subscriber growth is influenced more by population size and internet penetration
- Some high-unemployment countries still show strong growth
- India is the clear leader, while the US has high unemployment but moderate growth

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

Output:

```
Distribution of video views
for the last 30 days vary
across different channel
types:
 channel type
Animals           7.073477e+08
Entertainment     2.128728e+08
Sports            2.030792e+08
Education         2.007373e+08
Comedy            1.923643e+08
News              1.810388e+08
People            1.792531e+08
Music             1.788401e+08
Autos             1.766301e+08
Film              1.405741e+08
Games             1.044434e+08
Nonprofit         9.759050e+07
Howto             5.865822e+07
Tech              5.508947e+07
Name:
video views for the last_30_d
ays, dtype: float64
Correlation between
subscribers gained in 30 days
and unemployment rate:
-0.06294613050037096
```


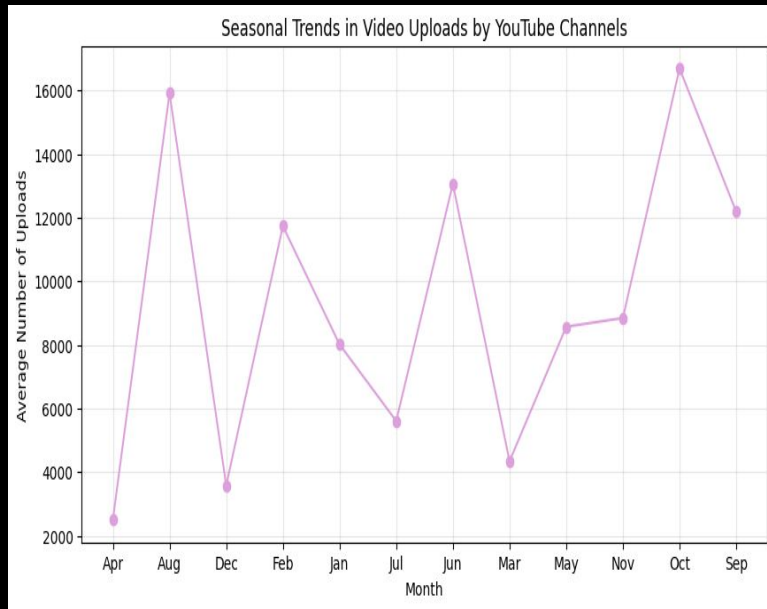Distribution of Video Views in Last 30 Days by Channel Type

Takeaways:

- Video views vary significantly by channel type, with Animals leading by a large margin
- Howto and Tech channels have the lowest total views
- **Correlation between subscribers gained and unemployment rate is very weak (-0.06)**

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

OUTPUT:

Seasonal Trends in Video Uploads by YouTube Channels

Seasonal trends
(sorted by uploads):

 created_month
Oct   16688.500000
Aug   15912.265060
Jun   13059.413333
Sep   12182.113402
Feb   11732.582090
Nov   8841.582418
May   8568.717647
Jan   8012.594059
Jul   5603.808989
Mar   4339.232558
Dec   3577.154930
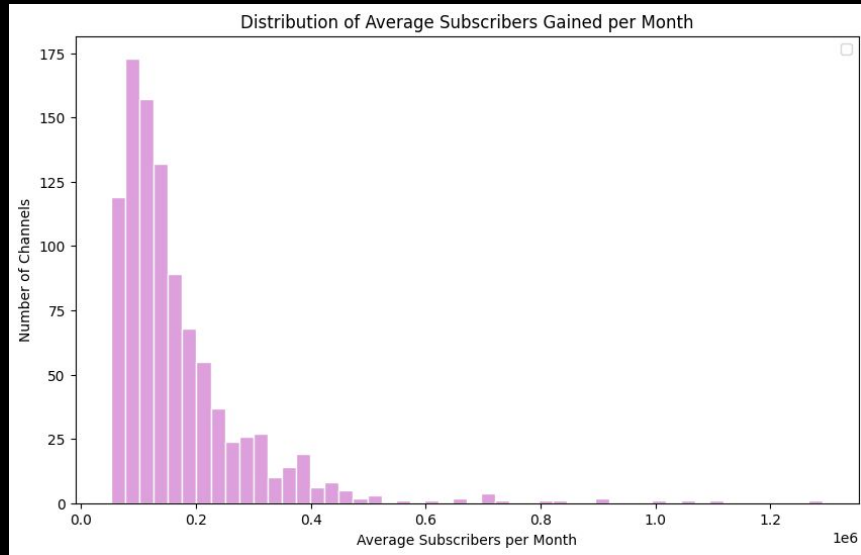Apr   2499.929577
Name: uploads, dtype: float64

TAKEAWAYS

- October and August have the highest average video uploads
- June, September, and February also show high activity
- April and December have the lowest video uploads
- Upload trends show peaks in late summer and fall. This may be because of the end of holiday season.

20. What is the average number of subscribers gained per month since the creation of YouTube

Output:

Distribution of Average Subscribers Gained per Month

```
Average number of subscribers gained per month
since channel creation: 171107.50708318167
```

Takeaways:

- Average subscribers gained per month is around 171,108.
- Most channels gain between 50k–200k subscribers per month
- Distribution is right-skewed with a few channels gaining 500k+.
- Majority of channels grow at a modest rate whereas a few dominate growth

THANK YOU