global-youtube-statistics-1

September 14, 2023

[2]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.graph_objects as go
```

[3]:
```
df=pd.read_excel(r"C:\Users\prera\Downloads\Global YouTube Statistics (1).xlsx")
```

[4]:
```
df
```

[4]:
```
     rank                    Youtuber  subscribers   video views  \
0       1                     T-Series    245000000  228000000000
1       2                YouTube Movies    170000000             0
2       3                      MrBeast    166000000   28368841870
3       4  Cocomelon - Nursery Rhymes    162000000  164000000000
4       5                    SET India    159000000  148000000000
..    …                            …           …            …
990   991                 Natan por Aï¿     12300000    9029609749
991   992    Free Fire India Official     12300000    1674409945
992   993                       Panda     12300000    2214684303
993   994                  RobTopGames     12300000     374123483
994   995                 Make Joke Of     12300000    2129773714

              category                        Title  uploads         Country  \
0                Music                      T-Series    20082           India
1    Film & Animation                youtubemovies        1   United States
2       Entertainment                      MrBeast      741   United States
3           Education  Cocomelon - Nursery Rhymes      966   United States
4               Shows                    SET India   116536           India
..                …                            …      …              …
990           Sports                 Natan por Aï¿     1200          Brazil
991    People & Blogs    Free Fire India Official     1500           India
992               NaN                  HybridPanda     2452  United Kingdom
993            Gaming                  RobTopGames       39          Sweden
994            Comedy                 Make Joke Of       62           India

    Abbreviation   channel_type  …  subscribers_for_last_30_days  \
```

```
0             IN          Music  …                    2000000.0
1             US          Games  …                          NaN
2             US  Entertainment  …                    8000000.0
3             US      Education  …                    1000000.0
4             IN  Entertainment  …                    1000000.0
..           …              …  …                          …
990           BR  Entertainment  …                     700000.0
991           IN          Games  …                     300000.0
992           GB          Games  …                       1000.0
993           SE          Games  …                     100000.0
994           IN         Comedy  …                     100000.0

     created_year  created_month  created_date  \
0          2006.0            Mar          13.0
1          2006.0            Mar           5.0
2          2012.0            Feb          20.0
3          2006.0            Sep           1.0
4          2006.0            Sep          20.0
..            …              …             …
990        2017.0            Feb          12.0
991        2018.0            Sep          14.0
992        2006.0            Sep          11.0
993        2012.0            May           9.0
994        2017.0            Aug           1.0

     Gross tertiary education enrollment (%)    Population  Unemployment rate  \
0                                       28.1  1.366418e+09               5.36
1                                       88.2  3.282395e+08              14.70
2                                       88.2  3.282395e+08              14.70
3                                       88.2  3.282395e+08              14.70
4                                       28.1  1.366418e+09               5.36
..                                       …           …                  …
990                                     51.3  2.125594e+08              12.08
991                                     28.1  1.366418e+09               5.36
992                                     60.0  6.683440e+07               3.85
993                                     67.0  1.028545e+07               6.48
994                                     28.1  1.366418e+09               5.36

     Urban_population   Latitude  Longitude
0         471031528.0  20.593684  78.962880
1         270663028.0  37.090240 -95.712891
2         270663028.0  37.090240 -95.712891
3         270663028.0  37.090240 -95.712891
4         471031528.0  20.593684  78.962880
..                …          …          …
990       183241641.0 -14.235004 -51.925280
991       471031528.0  20.593684  78.962880
```

```
992        55908316.0  55.378051  -3.435973
993         9021165.0  60.128161  18.643501
994       471031528.0  20.593684  78.962880

[995 rows x 28 columns]
```

[5]:
```
df.shape
```

[5]:
```
(995, 28)
```

[6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 995 entries, 0 to 994
Data columns (total 28 columns):
 #   Column                                  Non-Null Count  Dtype
---  ------                                  --------------  -----
 0   rank                                    995 non-null    int64
 1   Youtuber                                995 non-null    object
 2   subscribers                             995 non-null    int64
 3   video views                             995 non-null    int64
 4   category                                949 non-null    object
 5   Title                                   995 non-null    object
 6   uploads                                 995 non-null    int64
 7   Country                                 873 non-null    object
 8   Abbreviation                            873 non-null    object
 9   channel_type                            965 non-null    object
 10  video_views_rank                        994 non-null    float64
 11  country_rank                            879 non-null    float64
 12  channel_type_rank                       962 non-null    float64
 13  video_views_for_the_last_30_days        939 non-null    float64
 14  lowest_monthly_earnings                 995 non-null    float64
 15  highest_monthly_earnings                995 non-null    float64
 16  lowest_yearly_earnings                  995 non-null    float64
 17  highest_yearly_earnings                 995 non-null    float64
 18  subscribers_for_last_30_days            658 non-null    float64
 19  created_year                            990 non-null    float64
 20  created_month                           990 non-null    object
 21  created_date                            990 non-null    float64
 22  Gross tertiary education enrollment (%) 872 non-null    float64
 23  Population                              872 non-null    float64
 24  Unemployment rate                       872 non-null    float64
 25  Urban_population                        872 non-null    float64
 26  Latitude                                872 non-null    float64
 27  Longitude                               872 non-null    float64
dtypes: float64(17), int64(4), object(7)
memory usage: 217.8+ KB
```

[7]:
```
df.describe()
```

[7]:

	rank	subscribers	video views	uploads	video_views_rank
count	995.00000	9.950000e+02	9.950000e+02	995.000000	9.940000e+02
mean	498.00000	2.298241e+07	1.103954e+10	9187.125628	5.542489e+05
std	287.37606	1.752611e+07	1.411084e+10	34151.352254	1.362782e+06
min	1.00000	1.230000e+07	0.000000e+00	0.000000	1.000000e+00
25%	249.50000	1.450000e+07	4.288145e+09	194.500000	3.230000e+02
50%	498.00000	1.770000e+07	7.760820e+09	729.000000	9.155000e+02
75%	746.50000	2.460000e+07	1.355470e+10	2667.500000	3.584500e+03
max	995.00000	2.450000e+08	2.280000e+11	301308.000000	4.057944e+06

	country_rank	channel_type_rank	video_views_for_the_last_30_days
count	879.000000	962.000000	9.390000e+02
mean	386.053470	745.719335	1.756103e+08
std	1232.244746	1944.386561	4.163782e+08
min	1.000000	1.000000	1.000000e+00
25%	11.000000	27.000000	2.013750e+07
50%	51.000000	65.500000	6.408500e+07
75%	123.000000	139.750000	1.688265e+08
max	7741.000000	7741.000000	6.589000e+09

	lowest_monthly_earnings	highest_monthly_earnings	…
count	995.000000	9.950000e+02	…
mean	36886.148281	5.898078e+05	…
std	71858.724092	1.148622e+06	…
min	0.000000	0.000000e+00	…
25%	2700.000000	4.350000e+04	…
50%	13300.000000	2.127000e+05	…
75%	37900.000000	6.068000e+05	…
max	850900.000000	1.360000e+07	…

	highest_yearly_earnings	subscribers_for_last_30_days	created_year
count	9.950000e+02	6.580000e+02	990.000000
mean	7.081814e+06	3.490791e+05	2012.630303
std	1.379704e+07	6.143554e+05	4.512503
min	0.000000e+00	1.000000e+00	1970.000000
25%	5.217500e+05	1.000000e+05	2009.000000
50%	2.600000e+06	2.000000e+05	2013.000000
75%	7.300000e+06	4.000000e+05	2016.000000
max	1.634000e+08	8.000000e+06	2022.000000

	created_date	Gross tertiary education enrollment (%)	Population
count	990.000000	872.000000	8.720000e+02
mean	15.746465	63.627752	4.303873e+08
std	8.777520	26.106893	4.727947e+08
min	1.000000	7.600000	2.025060e+05

```
25%        8.000000                                36.300000  8.335541e+07
50%       16.000000                                68.000000  3.282395e+08
75%       23.000000                                88.200000  3.282395e+08
max       31.000000                               113.100000  1.397715e+09

       Unemployment rate  Urban_population    Latitude   Longitude
count         872.000000      8.720000e+02  872.000000  872.000000
mean            9.279278      2.242150e+08   26.632783  -14.128146
std             4.888354      1.546874e+08   20.560533   84.760809
min             0.750000      3.558800e+04  -38.416097 -172.104629
25%             5.270000      5.590832e+07   20.593684  -95.712891
50%             9.365000      2.706630e+08   37.090240  -51.925280
75%            14.700000      2.706630e+08   37.090240   78.962880
max            14.720000      8.429340e+08   61.924110  138.252924

[8 rows x 21 columns]
```

[8]:
```
df.head(3)
```

[8]:
```
   rank        Youtuber  subscribers   video views          category  \
0     1        T-Series    245000000  228000000000             Music
1     2  YouTube Movies    170000000             0  Film & Animation
2     3         MrBeast    166000000   28368841870     Entertainment

           Title  uploads        Country Abbreviation   channel_type  …  \
0       T-Series    20082          India           IN          Music  …
1  youtubemovies        1  United States           US          Games  …
2        MrBeast      741  United States           US  Entertainment  …

   subscribers_for_last_30_days  created_year  created_month  created_date  \
0                     2000000.0        2006.0            Mar          13.0
1                           NaN        2006.0            Mar           5.0
2                     8000000.0        2012.0            Feb          20.0

   Gross tertiary education enrollment (%)    Population  Unemployment rate  \
0                                     28.1  1.366418e+09               5.36
1                                     88.2  3.282395e+08              14.70
2                                     88.2  3.282395e+08              14.70

   Urban_population   Latitude  Longitude
0       471031528.0  20.593684  78.962880
1       270663028.0  37.090240 -95.712891
2       270663028.0  37.090240 -95.712891

[3 rows x 28 columns]
```

[9]:
```
df.tail(3)
```

[9]:
```
     rank      Youtuber  subscribers  video views category         Title  \
992   993         Panda     12300000   2214684303      NaN   HybridPanda
993   994   RobTopGames     12300000    374123483   Gaming   RobTopGames
994   995  Make Joke Of     12300000   2129773714   Comedy  Make Joke Of

     uploads         Country Abbreviation channel_type  …  \
992     2452  United Kingdom           GB        Games  …
993       39          Sweden           SE        Games  …
994       62           India           IN       Comedy  …

     subscribers_for_last_30_days  created_year  created_month  created_date  \
992                        1000.0        2006.0            Sep          11.0
993                      100000.0        2012.0            May           9.0
994                      100000.0        2017.0            Aug           1.0

     Gross tertiary education enrollment (%)    Population  Unemployment rate  \
992                                     60.0  6.683440e+07               3.85
993                                     67.0  1.028545e+07               6.48
994                                     28.1  1.366418e+09               5.36

     Urban_population   Latitude  Longitude
992       55908316.0  55.378051  -3.435973
993        9021165.0  60.128161  18.643501
994      471031528.0  20.593684  78.962880

[3 rows x 28 columns]
```

[10]:
```
df.nunique()
```

[10]:
```
rank                                995
Youtuber                            995
subscribers                         289
video views                         988
category                             18
Title                               992
uploads                             777
Country                              49
Abbreviation                         49
channel_type                         14
video_views_rank                    953
country_rank                        246
channel_type_rank                   286
video_views_for_the_last_30_days    908
lowest_monthly_earnings             557
highest_monthly_earnings            736
lowest_yearly_earnings              757
highest_yearly_earnings             419
```

```
subscribers_for_last_30_days                53
created_year                                19
created_month                               12
created_date                                31
Gross tertiary education enrollment (%)     47
Population                                  48
Unemployment rate                           47
Urban_population                            48
Latitude                                    48
Longitude                                   48
dtype: int64
```

[11]:
```
df.isnull().sum()
```

[11]:
```
rank                                         0
Youtuber                                     0
subscribers                                  0
video views                                  0
category                                    46
Title                                        0
uploads                                      0
Country                                    122
Abbreviation                               122
channel_type                                30
video_views_rank                             1
country_rank                               116
channel_type_rank                           33
video_views_for_the_last_30_days            56
lowest_monthly_earnings                      0
highest_monthly_earnings                     0
lowest_yearly_earnings                       0
highest_yearly_earnings                      0
subscribers_for_last_30_days               337
created_year                                 5
created_month                                5
created_date                                 5
Gross tertiary education enrollment (%)    123
Population                                 123
Unemployment rate                          123
Urban_population                           123
Latitude                                   123
Longitude                                  123
dtype: int64
```

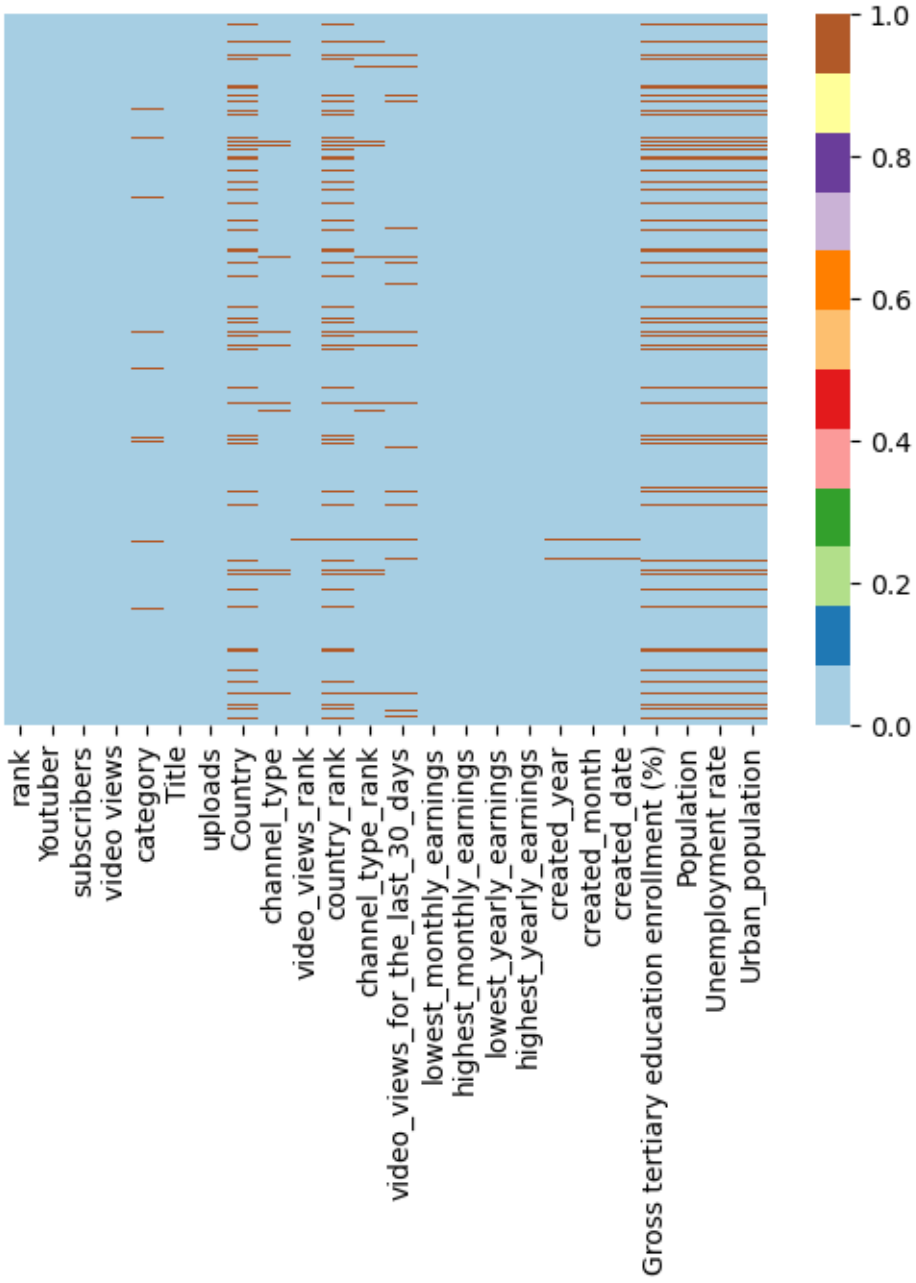
[12]:
```
(df.isnull().sum()/(len(df)))*100 #percentage of missing values in each column
```

[12]:
```
rank                                    0.000000
Youtuber                                0.000000
subscribers                             0.000000
video views                             0.000000
category                                4.623116
Title                                   0.000000
uploads                                 0.000000
Country                                12.261307
Abbreviation                           12.261307
channel_type                            3.015075
video_views_rank                        0.100503
country_rank                           11.658291
channel_type_rank                       3.316583
video_views_for_the_last_30_days        5.628141
lowest_monthly_earnings                 0.000000
highest_monthly_earnings                0.000000
lowest_yearly_earnings                  0.000000
highest_yearly_earnings                 0.000000
subscribers_for_last_30_days           33.869347
created_year                            0.502513
created_month                           0.502513
created_date                            0.502513
Gross tertiary education enrollment (%)    12.361809
Population                             12.361809
Unemployment rate                      12.361809
Urban_population                       12.361809
Latitude                               12.361809
Longitude                              12.361809
dtype: float64
```

[13]:
```
df_=df.
↪drop(["Longitude","Latitude","Abbreviation","subscribers_for_last_30_days"],axis=1,inplace='
```

[14]:
```
sns.heatmap(df.isnull(),yticklabels=False,cmap="Paired") #to check the missing␣
↪values on a heatmap
```

[14]:
```
<Axes: >
```



[15]:
```
df.dropna(subset=["category"],inplace=True)
```

[16]:
```
df.dropna(subset=["channel_type"],inplace=True)
df.dropna(subset=["channel_type_rank"],inplace=True)
df.dropna(subset=["video_views_for_the_last_30_days"],inplace=True)
df.dropna(subset=["created_year"],inplace=True)
df.dropna(subset=["created_month"],inplace=True)
df.dropna(subset=["created_date"],inplace=True)
```

[17]:
```
df.isnull().sum()
```

[17]:
```
rank                                    0
Youtuber                                0
subscribers                             0
video views                             0
category                                0
Title                                   0
uploads                                 0
Country                                80
channel_type                            0
video_views_rank                        0
country_rank                           74
channel_type_rank                       0
video_views_for_the_last_30_days        0
lowest_monthly_earnings                 0
highest_monthly_earnings                0
lowest_yearly_earnings                  0
highest_yearly_earnings                 0
created_year                            0
created_month                           0
created_date                            0
Gross tertiary education enrollment (%)    81
Population                              81
Unemployment rate                       81
Urban_population                        81
dtype: int64
```

[18]:
```
df["country_rank"].fillna(df["country_rank"].mean(),inplace=True)
```

[19]:
```
df["Country"].ffill(axis=0,inplace=True)
```

[20]:
```
df["Gross tertiary education enrollment (%)"].fillna(df["Gross tertiary␣
 ↪education enrollment (%)"].mean(),inplace=True)
df["Population"].fillna(df["Population"].mean(),inplace=True)
df["Unemployment rate"].fillna(df["Unemployment rate"].mean(),inplace=True)
df["Urban_population"].fillna(df["Urban_population"].mean(),inplace=True)
```

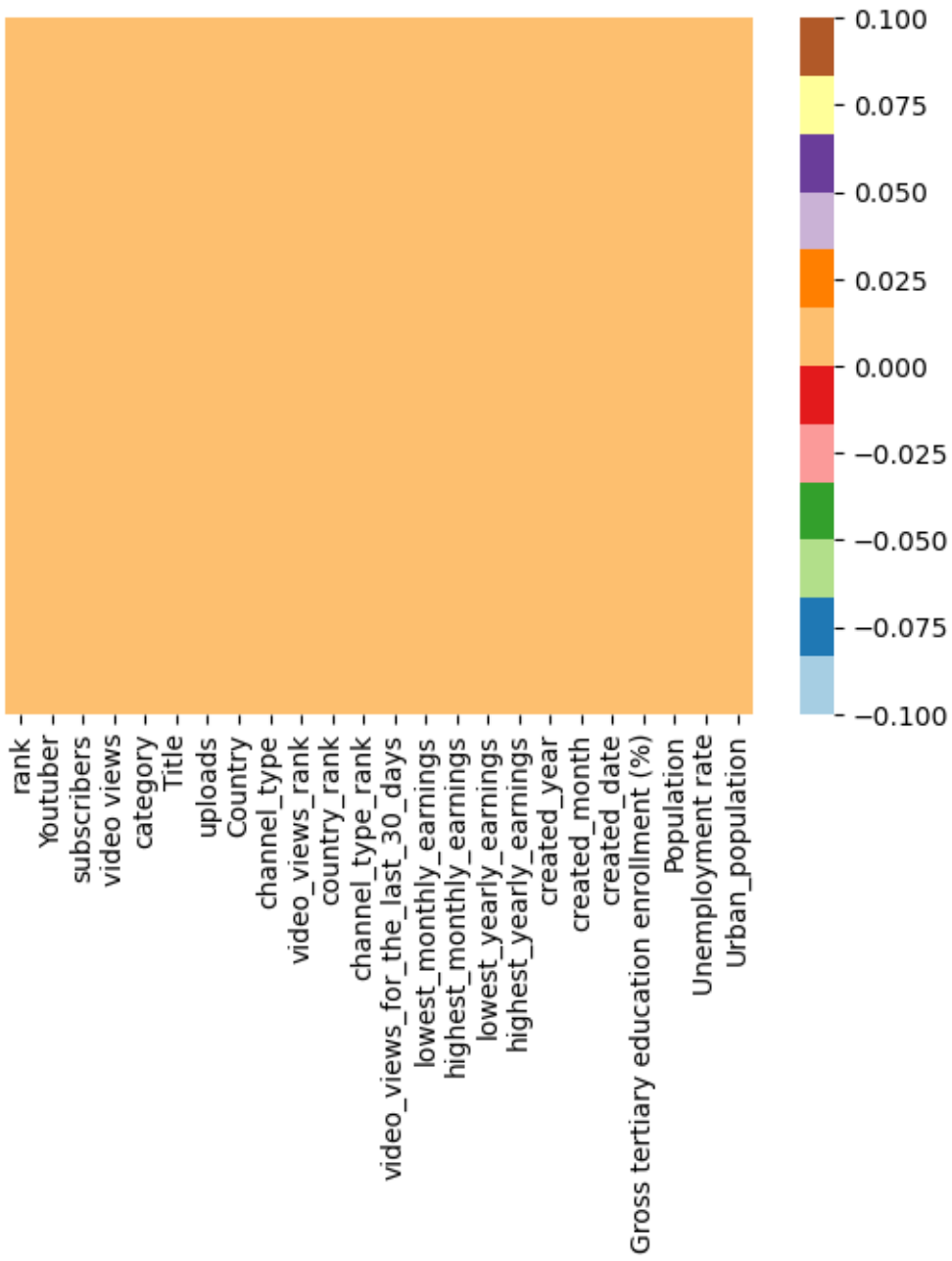
[21]:
```
df.isnull().sum()
```

[21]:
```
rank                    0
Youtuber                0
subscribers             0
video views             0
category                0
Title                   0
uploads                 0
```

```
Country                                    0
channel_type                               0
video_views_rank                           0
country_rank                               0
channel_type_rank                          0
video_views_for_the_last_30_days           0
lowest_monthly_earnings                    0
highest_monthly_earnings                   0
lowest_yearly_earnings                     0
highest_yearly_earnings                    0
created_year                               0
created_month                              0
created_date                               0
Gross tertiary education enrollment (%)    0
Population                                 0
Unemployment rate                          0
Urban_population                           0
dtype: int64
```

[22]:
```
sns.heatmap(df.isnull(),yticklabels=False,cmap="Paired")
```

[22]:
```
<Axes: >
```



[23]:
```
year_count = df.groupby('created_year')['Youtuber'].count().reset_index()␣
  ↪#number of youtubers in specific year
year_count
```

[23]:

	created_year	Youtuber
0	1970.0	1
1	2005.0	22
2	2006.0	80

```
3          2007.0         46
4          2008.0         42
5          2009.0         51
6          2010.0         44
7          2011.0         80
8          2012.0         63
9          2013.0         68
10         2014.0         97
11         2015.0         72
12         2016.0         71
13         2017.0         58
14         2018.0         34
15         2019.0         20
16         2020.0         24
17         2021.0         15
18         2022.0          1
```

[24]:
```
top15Youtubers =  df[['Youtuber','subscribers','category',
                      'created_year','Country']].head(15)   # Top 15 youtubers␣
 ↪from the dataset
top15Youtubers
```

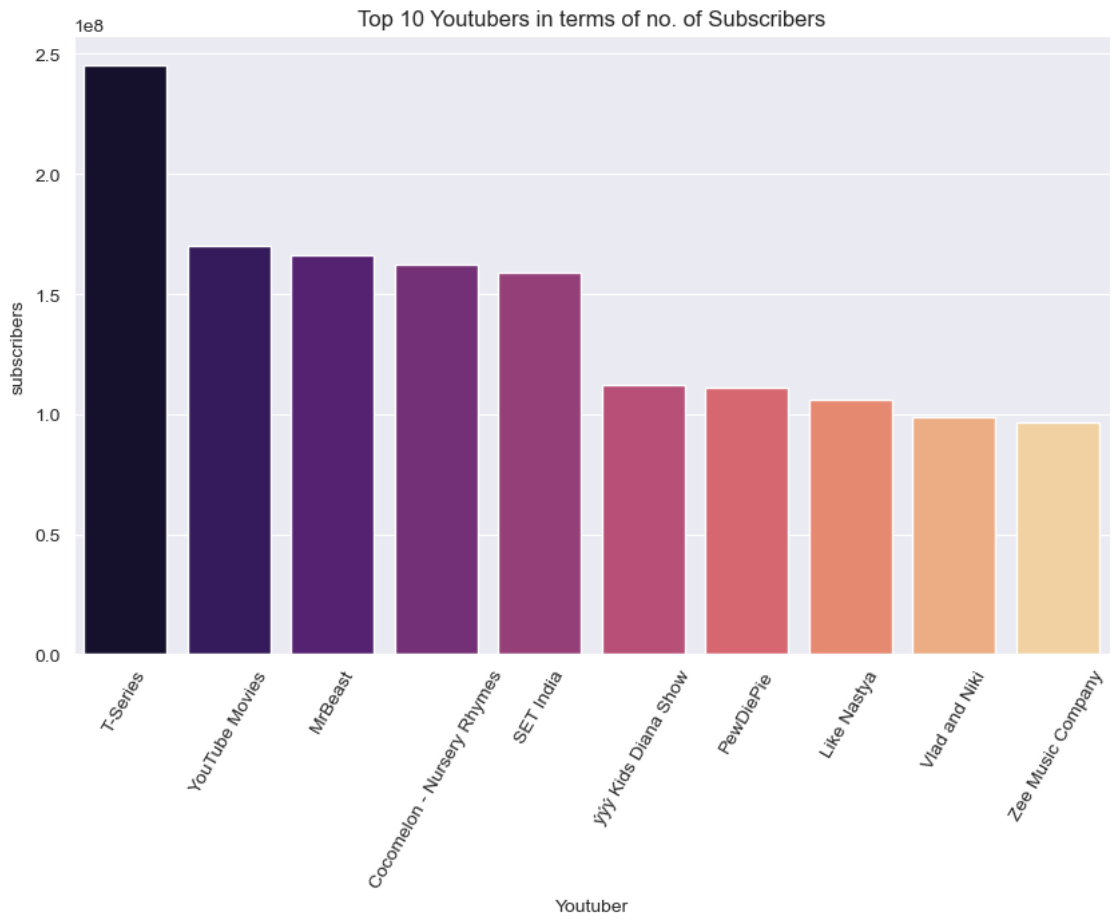
[24]:
```
                     Youtuber  subscribers          category  created_year  \
0                     T-Series    245000000             Music        2006.0
1                YouTube Movies    170000000  Film & Animation        2006.0
2                       MrBeast    166000000     Entertainment        2012.0
3   Cocomelon - Nursery Rhymes    162000000         Education        2006.0
4                     SET India    159000000             Shows        2006.0
6          ýýý Kids Diana Show    112000000    People & Blogs        2015.0
7                     PewDiePie    111000000            Gaming        2010.0
8                   Like Nastya    106000000    People & Blogs        2016.0
9                 Vlad and Niki     98900000     Entertainment        2018.0
10           Zee Music Company     96700000             Music        2014.0
11                         WWE     96000000            Sports        2007.0
13                   BLACKPINK     89800000    People & Blogs        2016.0
14                   Goldmines     86900000  Film & Animation        2006.0
15                   Sony SAB      83000000             Shows        2007.0
16             5-Minute Crafts     80100000     Howto & Style        2020.0

           Country
0            India
1    United States
2    United States
3    United States
4            India
6    United States
7            Japan
```

```
8            Russia
9     United States
10            India
11    United States
13      South Korea
14      South Korea
15            India
16   United Kingdom
```

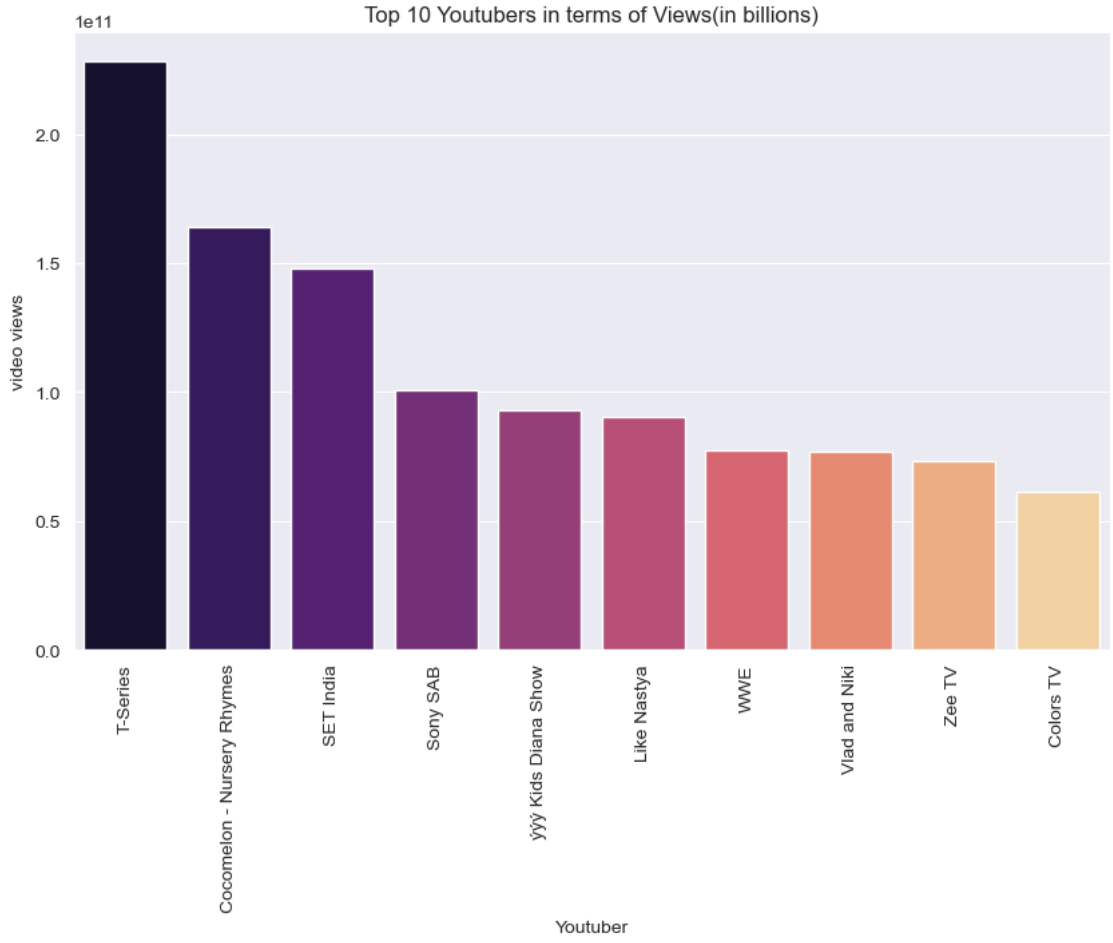
[25]:
```
sns.set_style('darkgrid')
plt.figure(figsize=(10,6))
ax=sns.barplot(y='subscribers',x='Youtuber',data=df,order=df.
 ↪sort_values('subscribers',ascending=False).Youtuber.iloc[:
 ↪10],palette='magma')
plt.xticks(rotation=60)
plt.title('Top 10 Youtubers in terms of no. of Subscribers')
```

[25]:
```
Text(0.5, 1.0, 'Top 10 Youtubers in terms of no. of Subscribers')
```



[26]:
```
sns.set_style('darkgrid')
plt.figure(figsize=(10,6))
ax=sns.barplot(y='video views',x='Youtuber',data=df,order=df.sort_values('video␣
 ↪views',ascending=False).Youtuber.iloc[:10],palette='magma')
plt.xticks(rotation=90)
plt.title('Top 10 Youtubers in terms of Views(in billions)')
```

[26]:
```
Text(0.5, 1.0, 'Top 10 Youtubers in terms of Views(in billions)')
```



[27]:
```
#to seperate categorical values from the numerical values
categorical_features=[feature for feature in df.columns if df[feature].
 ↪dtype=='object']
numerical_features=[feature for feature in df.columns if feature not in␣
 ↪categorical_features]
df_num=df[numerical_features]
df_cat=df[categorical_features]
```

[28]:
```
df_num
```

[28]:

	rank	subscribers	video views	uploads	video_views_rank	country_rank	\
0	1	245000000	228000000000	20082	1.0	1.0	
1	2	170000000	0	1	4055159.0	7670.0	
2	3	166000000	28368841870	741	48.0	1.0	
3	4	162000000	164000000000	966	2.0	2.0	
4	5	159000000	148000000000	116536	3.0	2.0	
..	…	…	…	…	…	…	
989	990	12400000	6993406259	99	833.0	175.0	
990	991	12300000	9029609749	1200	525.0	55.0	
991	992	12300000	1674409945	1500	6141.0	125.0	
993	994	12300000	374123483	39	35112.0	4.0	
994	995	12300000	2129773714	62	4568.0	125.0	

	channel_type_rank	video_views_for_the_last_30_days	\
0	1.0	2.258000e+09	
1	7423.0	1.200000e+01	
2	1.0	1.348000e+09	
3	1.0	1.975000e+09	
4	2.0	1.824000e+09	
..	…	…	
989	171.0	4.941200e+07	
990	172.0	5.525130e+08	
991	69.0	6.473500e+07	
993	69.0	3.871000e+06	
994	44.0	2.400000e+07	

	lowest_monthly_earnings	highest_monthly_earnings	\
0	564600.0	9000000.00	
1	0.0	0.05	
2	337000.0	5400000.00	
3	493800.0	7900000.00	
4	455900.0	7300000.00	
..	…	…	
989	12400.0	197600.00	
990	138100.0	2200000.00	
991	16200.0	258900.00	
993	968.0	15500.00	
994	6000.0	96000.00	

	lowest_yearly_earnings	highest_yearly_earnings	created_year	\
0	6800000.00	1.084000e+08	2006.0	
1	0.04	5.800000e-01	2006.0	
2	4000000.00	6.470000e+07	2012.0	
3	5900000.00	9.480000e+07	2006.0	
4	5500000.00	8.750000e+07	2006.0	

```
..                    …                   …          …
989               148200.00        2.400000e+06     2012.0
990              1700000.00        2.650000e+07     2017.0
991               194200.00        3.100000e+06     2018.0
993                11600.00        1.858000e+05     2012.0
994                72000.00        1.200000e+06     2017.0

     created_date  Gross tertiary education enrollment (%)    Population  \
0            13.0                                    28.1  1.366418e+09
1             5.0                                    88.2  3.282395e+08
2            20.0                                    88.2  3.282395e+08
3             1.0                                    88.2  3.282395e+08
4            20.0                                    28.1  1.366418e+09
..            …                                       …           …
989          17.0                                    88.2  3.282395e+08
990          12.0                                    51.3  2.125594e+08
991          14.0                                    28.1  1.366418e+09
993           9.0                                    67.0  1.028545e+07
994           1.0                                    28.1  1.366418e+09

     Unemployment rate  Urban_population
0                 5.36       471031528.0
1                14.70       270663028.0
2                14.70       270663028.0
3                14.70       270663028.0
4                 5.36       471031528.0
..                 …                …
989              14.70       270663028.0
990              12.08       183241641.0
991               5.36       471031528.0
993               6.48         9021165.0
994               5.36       471031528.0

[889 rows x 18 columns]
```

[29]:
```
df_cat
```

[29]:
```
                      Youtuber          category                      Title  \
0                     T-Series             Music                    T-Series
1                YouTube Movies  Film & Animation                youtubemovies
2                      MrBeast     Entertainment                      MrBeast
3    Cocomelon - Nursery Rhymes         Education  Cocomelon - Nursery Rhymes
4                     SET India             Shows                     SET India
..                          …                 …                          …
989                  Migos ATL             Music                    Migos ATL
990               Natan por Aï¿            Sports                 Natan por Aï¿
991    Free Fire India Official    People & Blogs    Free Fire India Official
```

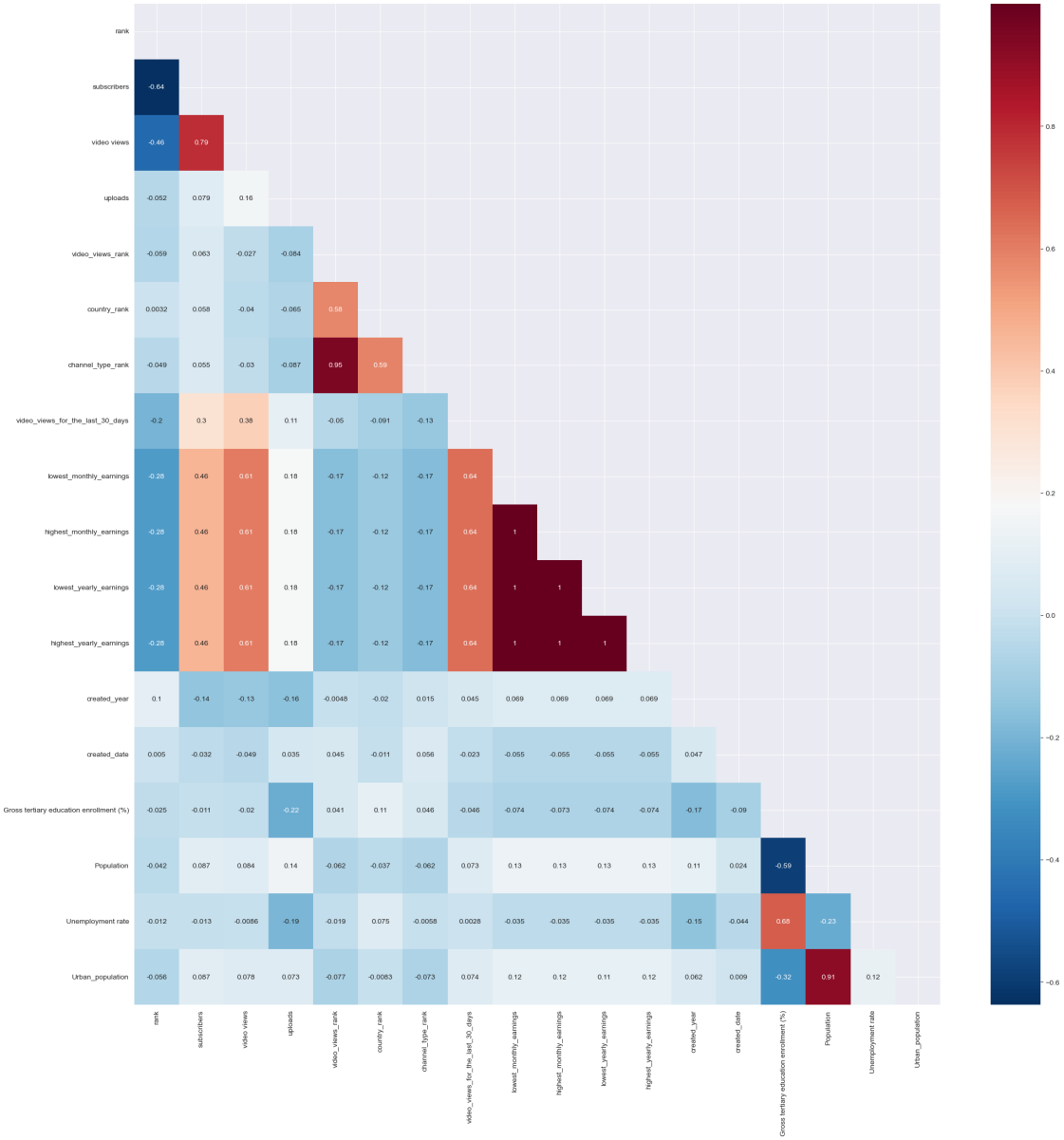
```
993                RobTopGames             Gaming                RobTopGames
994               Make Joke Of             Comedy               Make Joke Of

           Country   channel_type created_month
0            India          Music           Mar
1    United States          Games           Mar
2    United States  Entertainment           Feb
3    United States      Education           Sep
4            India  Entertainment           Sep
..             …              …             …
989  United States  Entertainment           Jan
990         Brazil  Entertainment           Feb
991          India          Games           Sep
993         Sweden          Games           May
994          India         Comedy           Aug

[889 rows x 6 columns]
```

[30]:
```
plt.figure(figsize=(25,25))
corr=df_num.corr()
msk = np.triu(np.ones_like(corr))
sns.heatmap(corr,cmap=plt.cm.RdBu_r,annot=True,annot_kws={'size':10},mask=msk)
```

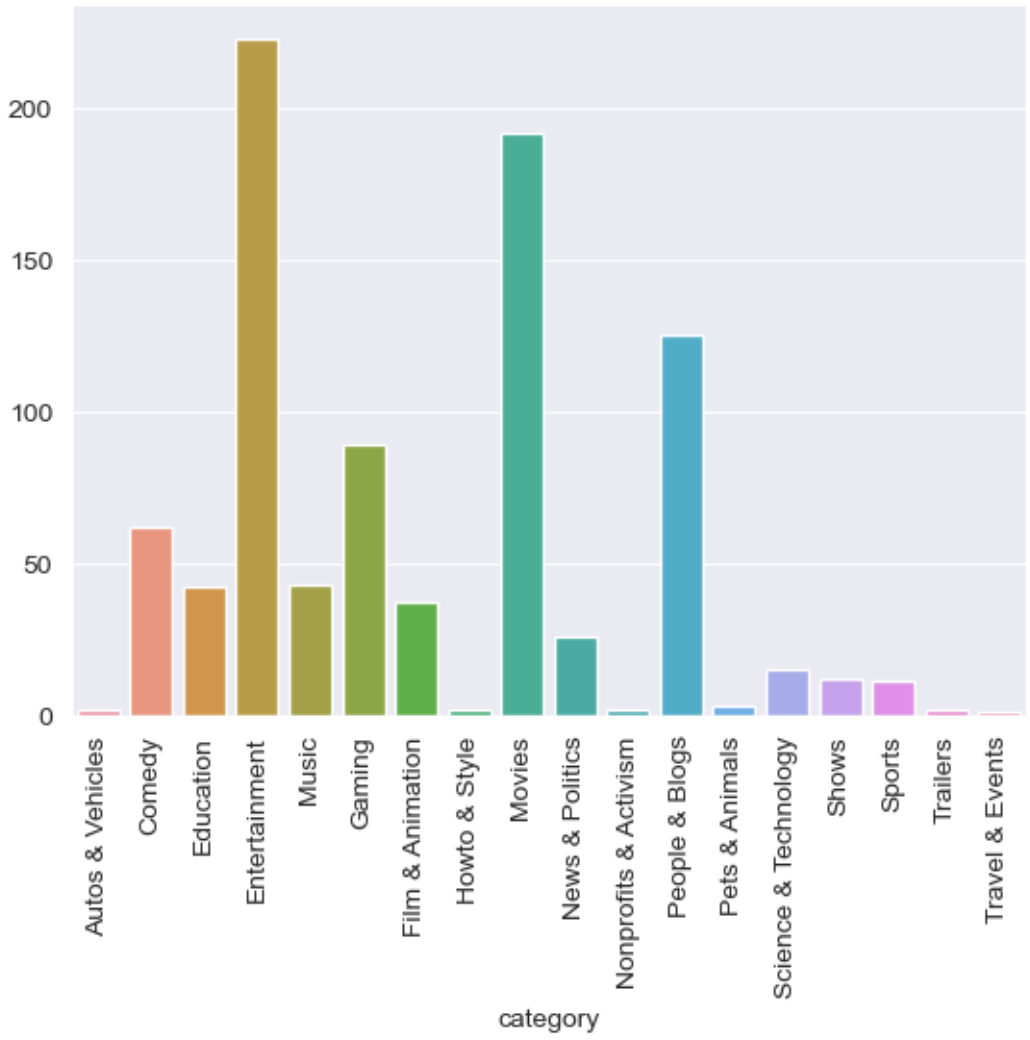
[30]:
```
<Axes: >
```



[31]:
```
c=df.groupby("category")["category"].count()
```

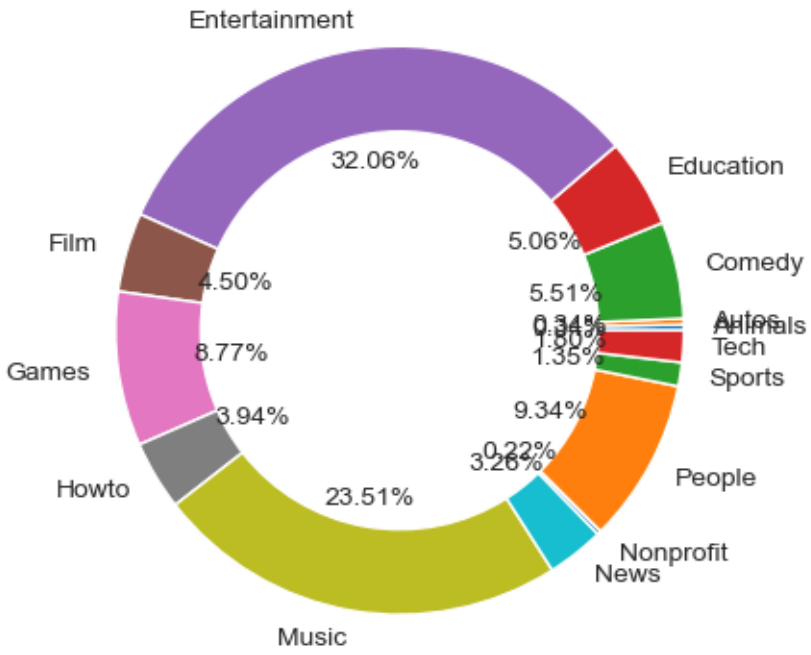
[32]:
```
a=sns.barplot(x=c.index, y=c.values)
a.set_xticklabels(
    labels=["Autos &␣
 ↪Vehicles","Comedy","Education","Entertainment","Music","Gaming","Film &␣
 ↪Animation","Howto & Style","Movies","News & Politics","Nonprofits &␣
 ↪Activism","People & Blogs","Pets & Animals","Science &␣
 ↪Technology","Shows","Sports","Trailers","Travel & Events"], rotation=90)
```

[32]: [Text(0, 0, 'Autos & Vehicles'),
 Text(1, 0, 'Comedy'),
 Text(2, 0, 'Education'),
 Text(3, 0, 'Entertainment'),
 Text(4, 0, 'Music'),
 Text(5, 0, 'Gaming'),
 Text(6, 0, 'Film & Animation'),
 Text(7, 0, 'Howto & Style'),
 Text(8, 0, 'Movies'),
 Text(9, 0, 'News & Politics'),
 Text(10, 0, 'Nonprofits & Activism'),
 Text(11, 0, 'People & Blogs'),
 Text(12, 0, 'Pets & Animals'),
 Text(13, 0, 'Science & Technology'),
 Text(14, 0, 'Shows'),
 Text(15, 0, 'Sports'),
 Text(16, 0, 'Trailers'),
 Text(17, 0, 'Travel & Events')]



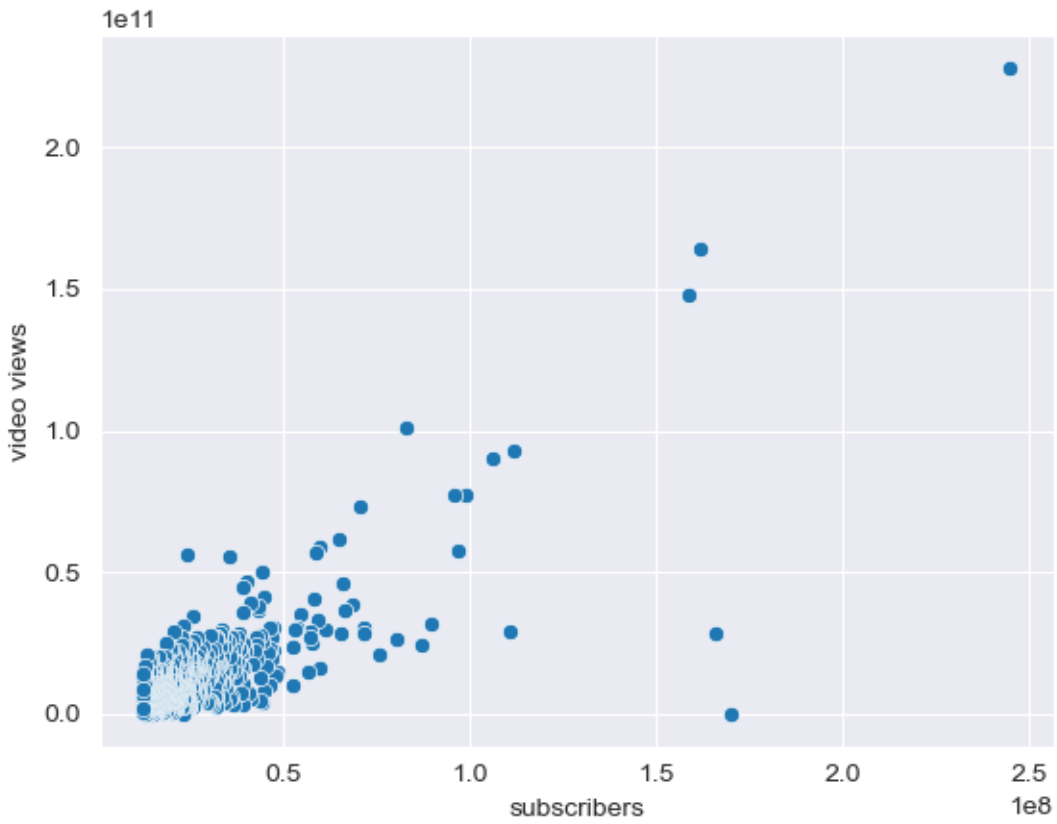
[33]:
```
c=df.groupby("channel_type")["channel_type"].count()
```

[34]:
```
plt.pie(c,labels=c.index,autopct="%.2f%%",wedgeprops=dict(width=0.3))
plt.show()
```



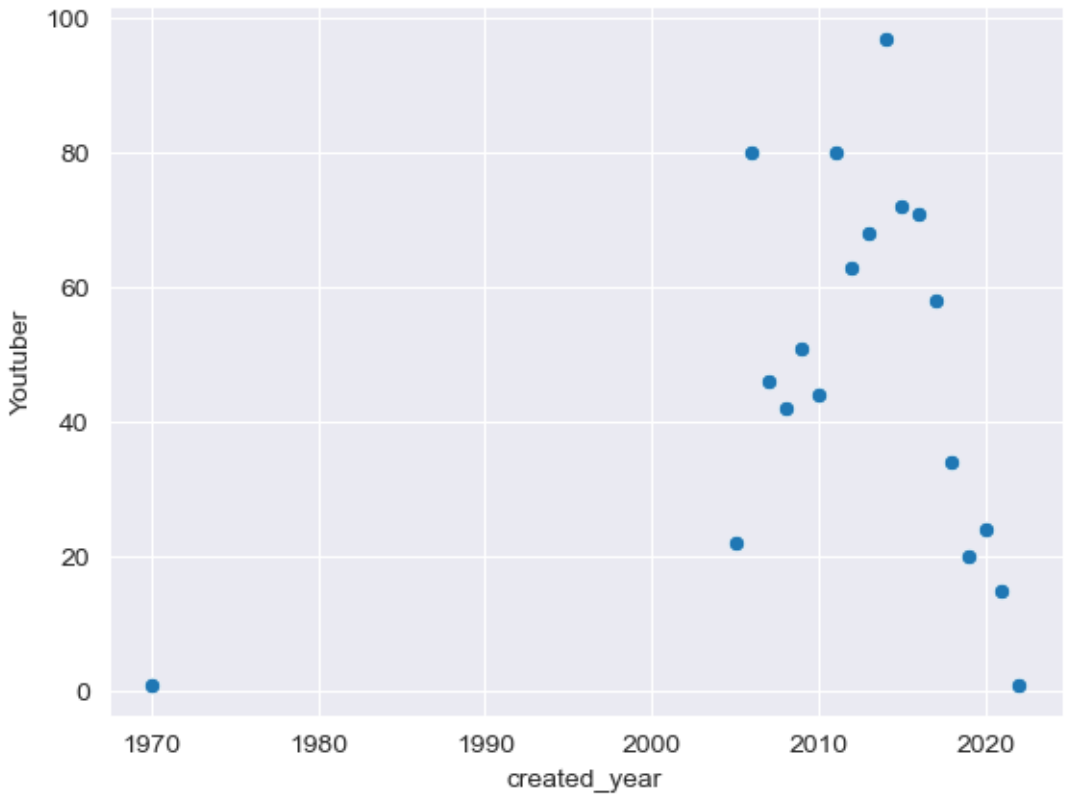
[35]:
```
sns.scatterplot(x="subscribers",y="video views",data=df)
```

[35]: <Axes: xlabel='subscribers', ylabel='video views'>

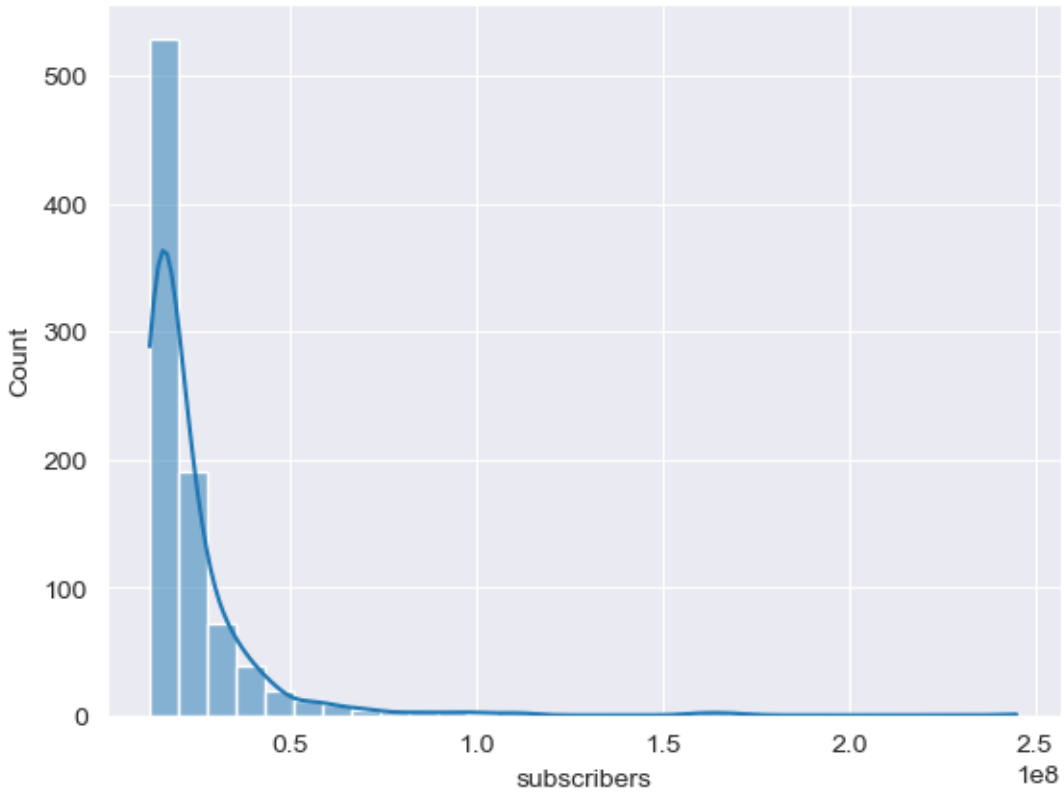


[36]:
```
sns.scatterplot(year_count , x = 'created_year' , y = 'Youtuber')
```

[36]: <Axes: xlabel='created_year', ylabel='Youtuber'>

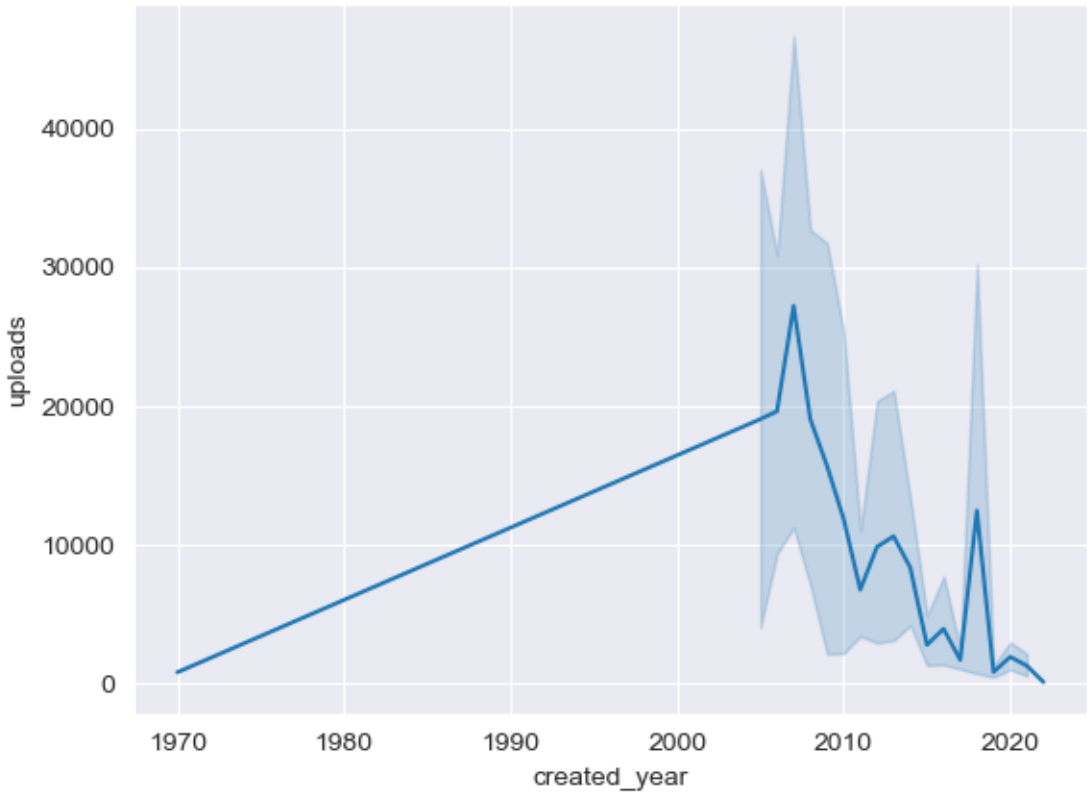


[37]:
```
sns.histplot(df['subscribers'], bins=30, kde=True)
plt.show()
```



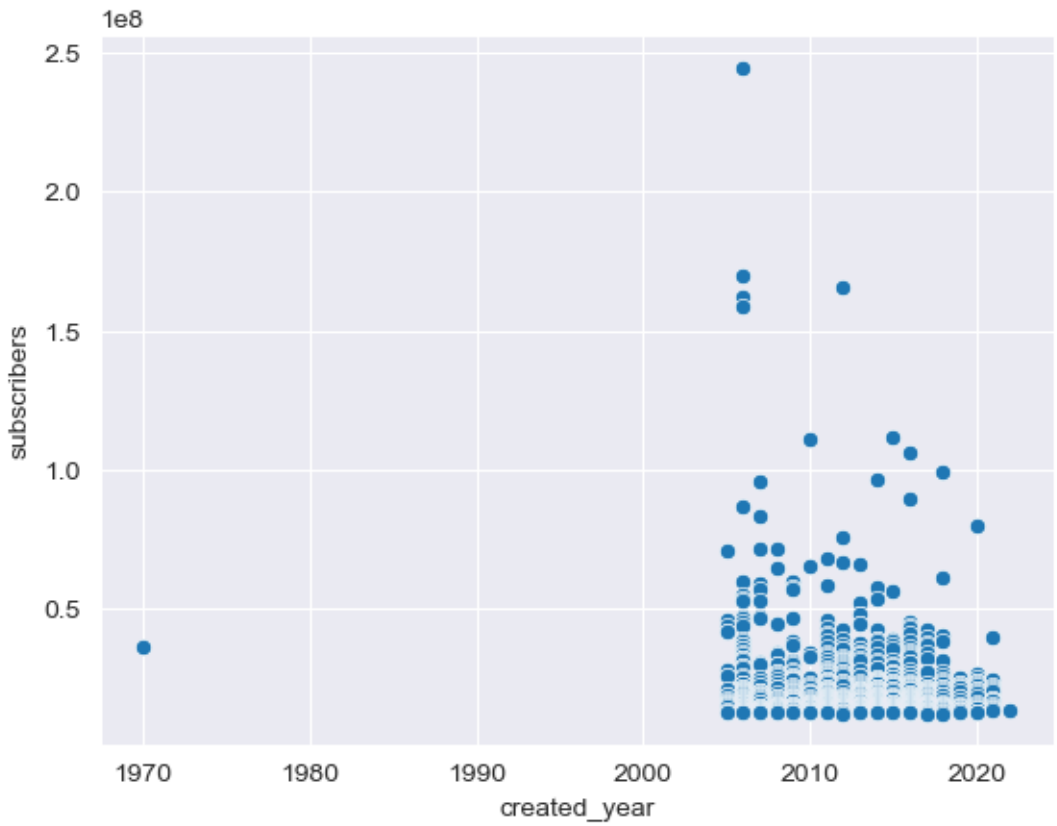
[38]:
```
sns.lineplot(x="created_year",y="uploads",data=df)
```

[38]: <Axes: xlabel='created_year', ylabel='uploads'>



[39]:
```
sns.scatterplot(x="created_year",y="subscribers",data=df)
```

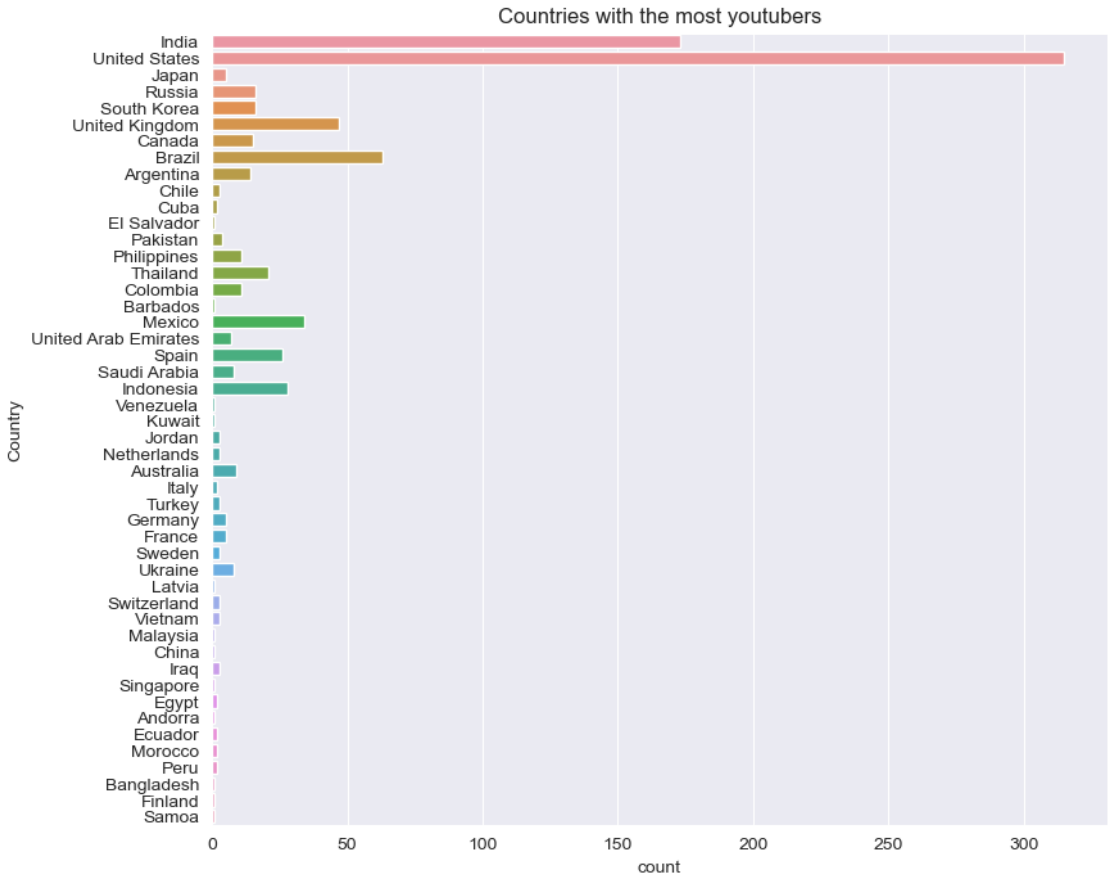
[39]:
```
<Axes: xlabel='created_year', ylabel='subscribers'>
```



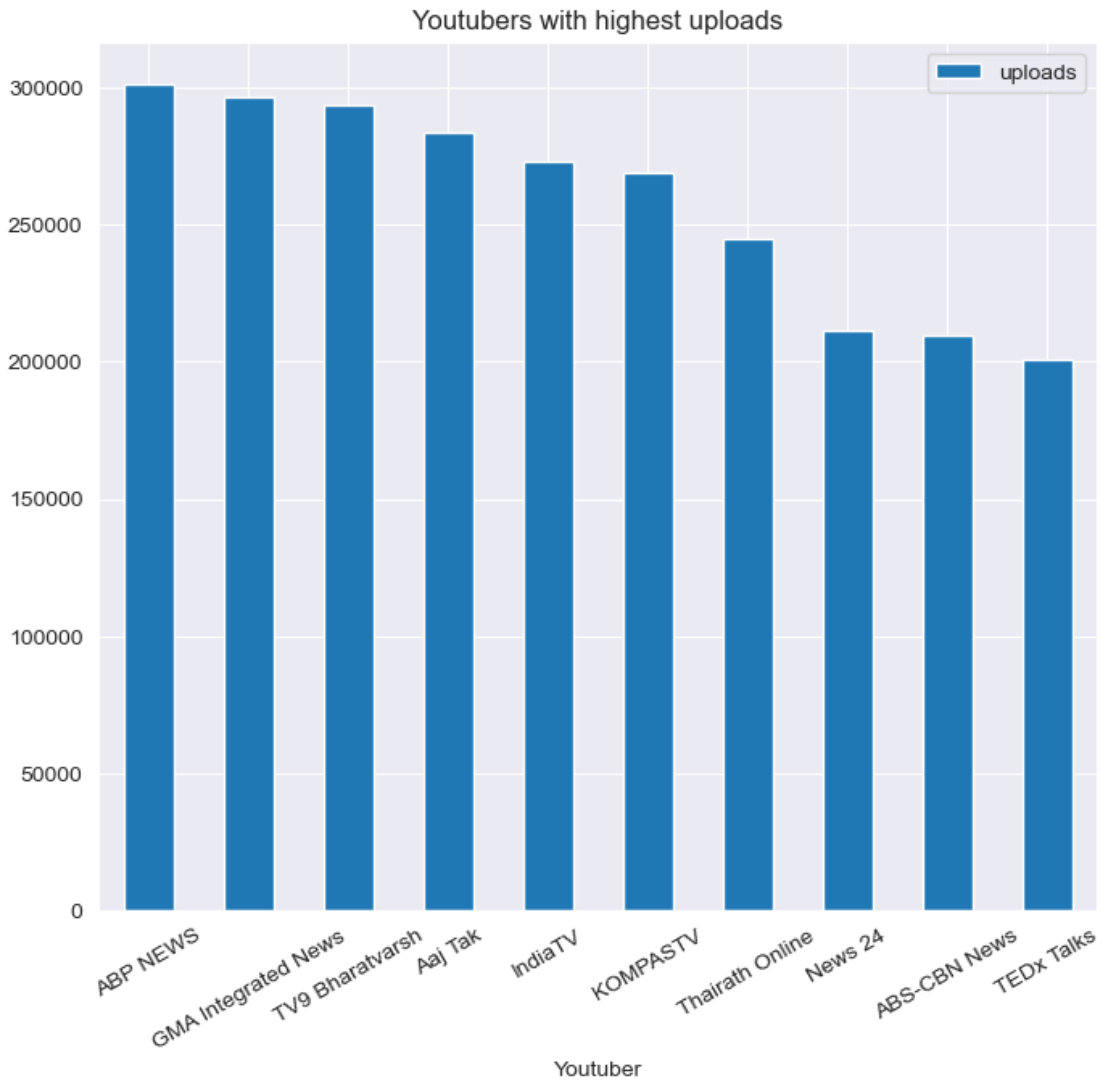
[40]:
```
country_counts = df['Country'].value_counts()[:10]
fig = go.Figure(go.Funnel(
    y=country_counts.index,
    x=country_counts.values,
    textinfo='value',
))

fig.update_layout(
    title_text='Top 10 Country Distribution',
    template='plotly_white'
)
fig.show()
```

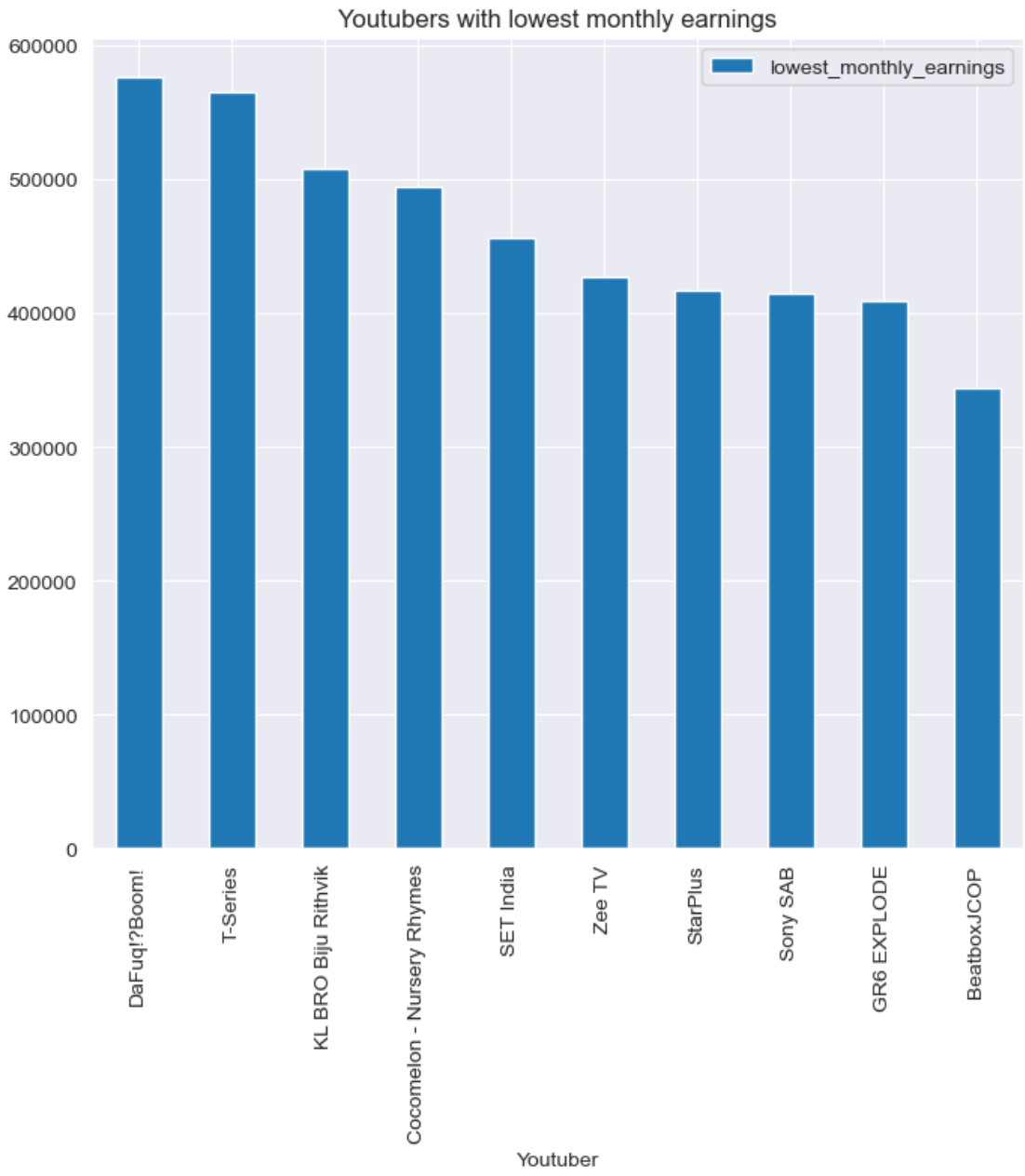
[41]:
```
sns.countplot(y=df['Country'])
plt.gcf().set_size_inches(9, 8)
plt.title('Countries with the most youtubers')
plt.show()
```



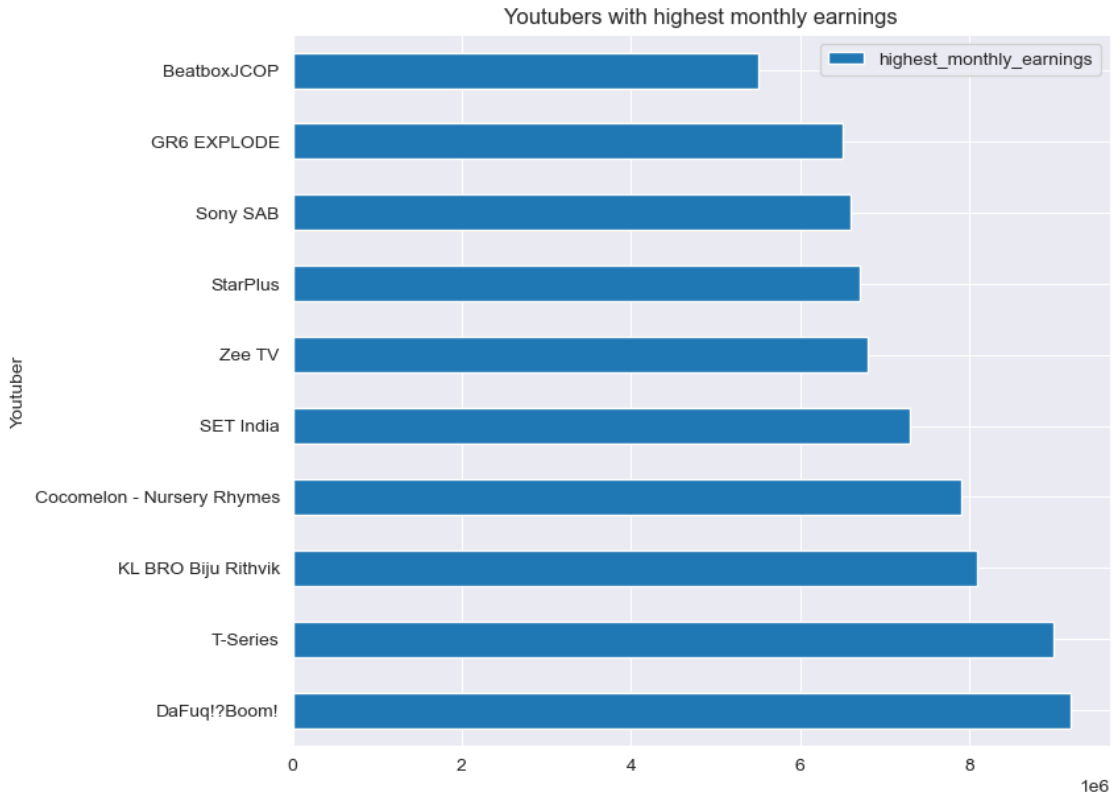
[42]:
```
df[['Youtuber', 'uploads']].nlargest(10, 'uploads').plot(kind='bar',␣
↪x='Youtuber', y='uploads', figsize=(8, 7), rot=30, title='Youtubers with␣
↪highest uploads')
plt.show()
```



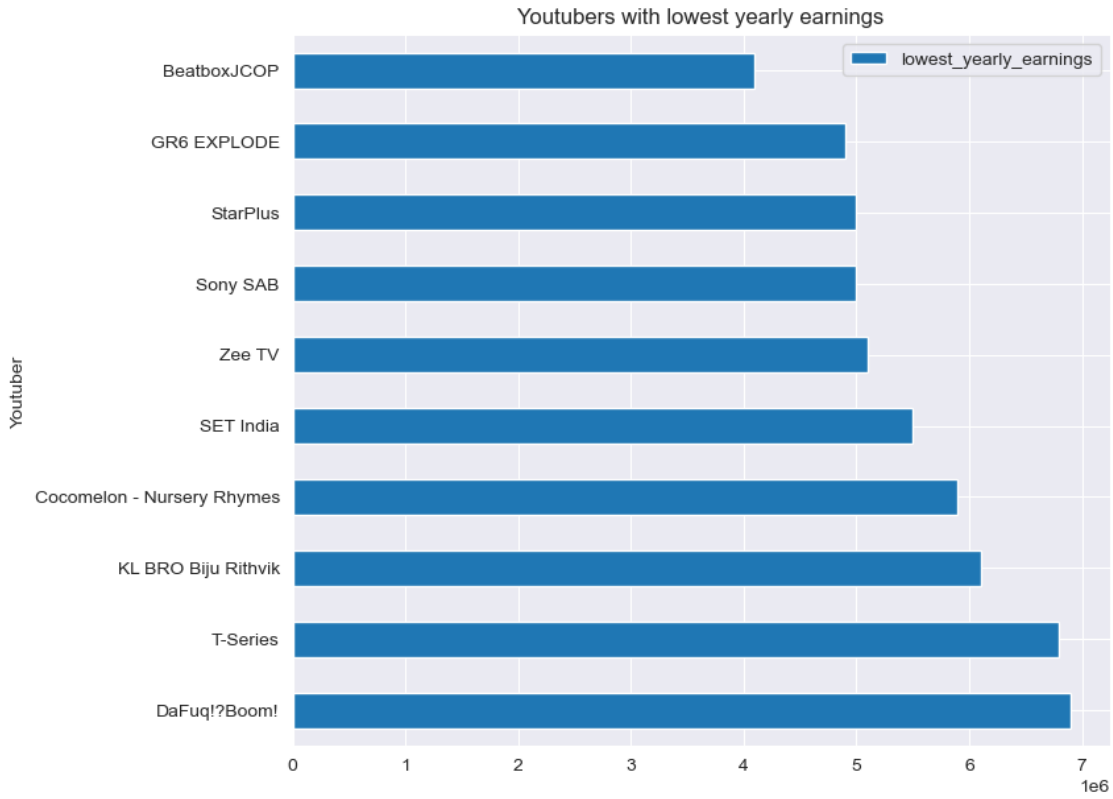
[43]:
```
df[['Youtuber', 'lowest_monthly_earnings']].nlargest(10,␣
 ↪'lowest_monthly_earnings').plot(kind='bar', x='Youtuber',␣
 ↪y='lowest_monthly_earnings', figsize=(8,7),
title='Youtubers with lowest monthly earnings');
```



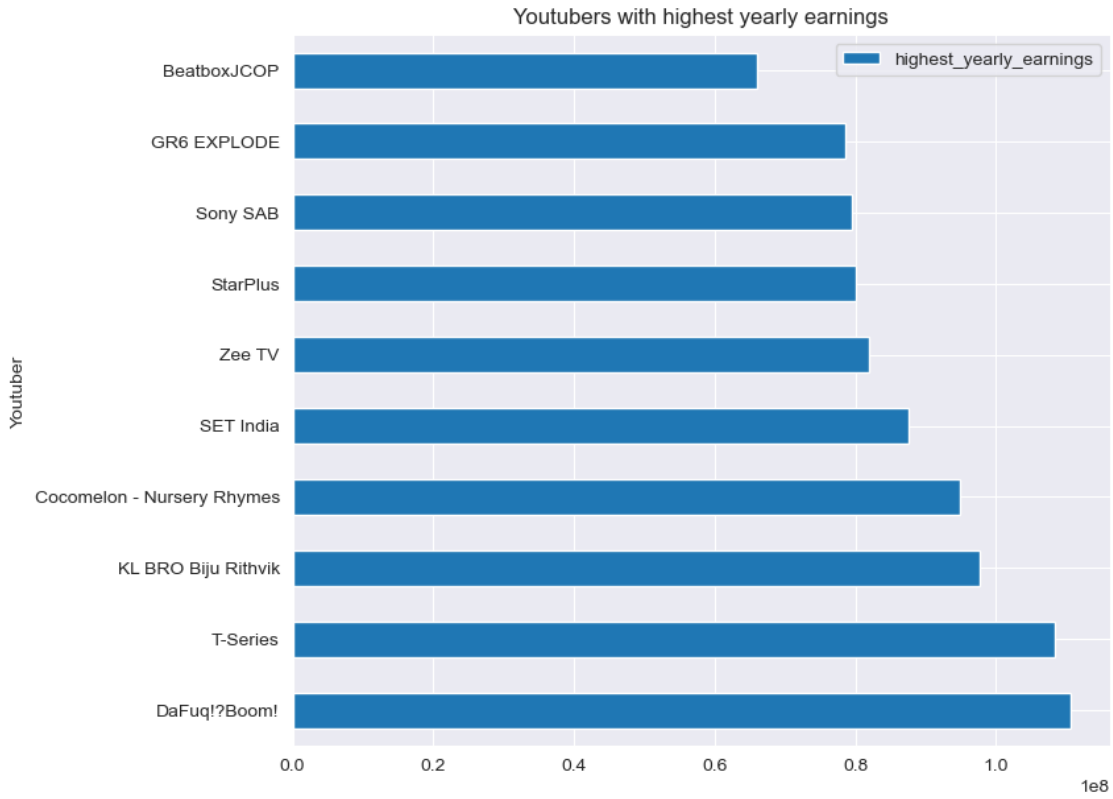
[44]:
```
df[['Youtuber', 'highest_monthly_earnings']].nlargest(10,␣
  ↪'highest_monthly_earnings').plot(kind='barh', x='Youtuber',␣
  ↪y='highest_monthly_earnings', figsize=(8,7),
title='Youtubers with highest monthly earnings');
```



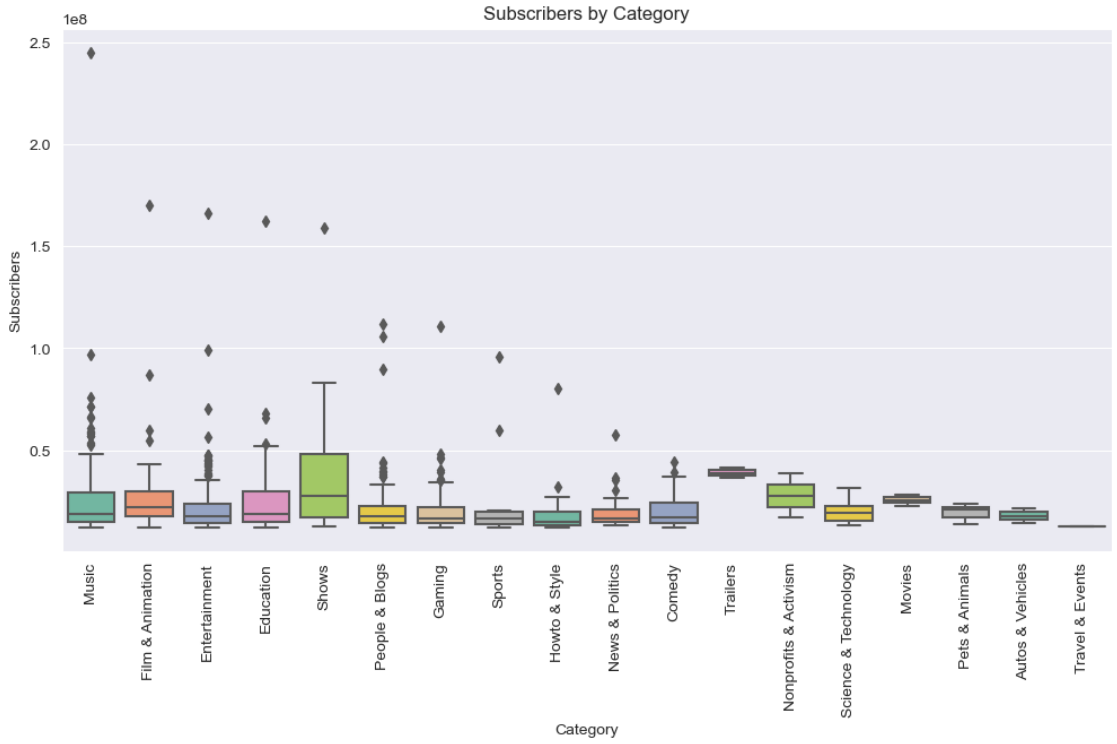
[45]:
```
df[['Youtuber', 'lowest_yearly_earnings']].nlargest(10,␣
↪'lowest_yearly_earnings').plot(kind='barh', x='Youtuber',␣
↪y='lowest_yearly_earnings', figsize=(8,7),
title='Youtubers with lowest yearly earnings');
```



[46]:
```
df[['Youtuber', 'highest_yearly_earnings']].nlargest(10,␣
↪'highest_yearly_earnings').plot(kind='barh', x='Youtuber',␣
↪y='highest_yearly_earnings', figsize=(8,7),
title='Youtubers with highest yearly earnings');
```



[47]:
```
plt.figure(figsize=(12, 6))
sns.boxplot(x='category', y='subscribers', data=df, palette='Set2')
plt.xlabel('Category')
plt.ylabel('Subscribers')
plt.title('Subscribers by Category')
plt.xticks(rotation=90)
plt.show()
```


Subscribers by Category
1e8
2.5
2.0
1.5
1.0
0.5
Subscribers
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

[ ]: