

```
import math
import warnings
import numpy as np
import pandas as pd
import seaborn as sns
import plotly.offline as py
import plotly.graph_objs as go
import matplotlib.pyplot as plt
```

```
yt=pd.read_csv('youtubers_df.csv')
yt.head(10)
```



	Rank	Username	Categories	Suscribers	Country	Visits	Likes	Com
0	1	tseries	Música y baile	249500000.0	India	86200.0	2700.0	
1	2	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18
2	3	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	
3	4	SETIndia	NaN	162600000.0	India	15600.0	166.0	
4	5	KidsDianaShow	Animación, Juguetes	113500000.0	Unknown	3900000.0	12400.0	
5	6	PewDiePie	Películas, Videojuegos	111500000.0	Estados Unidos	2400000.0	197300.0	4
6	7	LikeNastyaoofficial	Juguetes	107500000.0	Unknown	2600000.0	28000.0	
7	8	VladandNiki	Animación, Juguetes	101400000.0	Unknown	4100000.0	22100.0	
8	9	zeemusiccompany	Música y baile	99700000.0	India	74300.0	2600.0	
9	10	WWE	Videojuegos	97200000.0	Estados Unidos	184500.0	6300.0	

Next steps:

[Generate code with yt](#)

[View recommended plots](#)

yt.columns



```
Index(['Rank', 'Username', 'Categories', 'Suscribers', 'Country', 'Visits',
      'Likes', 'Comments', 'Links'],
      dtype='object')
```

yt.isnull().sum()



```
Rank      0
Username  0
Categories 306
Suscribers 0
Country   0
Visits    0
Likes     0
Comments  0
```

```
Links          0
dtype: int64
```

```
yt.info()
```

```
>>> <class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   Rank        1000 non-null   int64
 1   Username    1000 non-null   object
 2   Categories   694 non-null    object
 3   Suscribers  1000 non-null   float64
 4   Country     1000 non-null   object
 5   Visits      1000 non-null   float64
 6   Likes       1000 non-null   float64
 7   Comments    1000 non-null   float64
 8   Links       1000 non-null   object
dtypes: float64(4), int64(1), object(4)
memory usage: 70.4+ KB
```

```
yt["Categories"].info()
```

```
>>> <class 'pandas.core.series.Series'>
RangeIndex: 1000 entries, 0 to 999
Series name: Categories
Non-Null Count  Dtype
-----
694 non-null    object
dtypes: object(1)
memory usage: 7.9+ KB
```

```
yt['Country'].info()
```

```
>>> <class 'pandas.core.series.Series'>
RangeIndex: 1000 entries, 0 to 999
Series name: Country
Non-Null Count  Dtype
-----
1000 non-null   object
dtypes: object(1)
memory usage: 7.9+ KB
```

```
yt['Country'].mode()
```

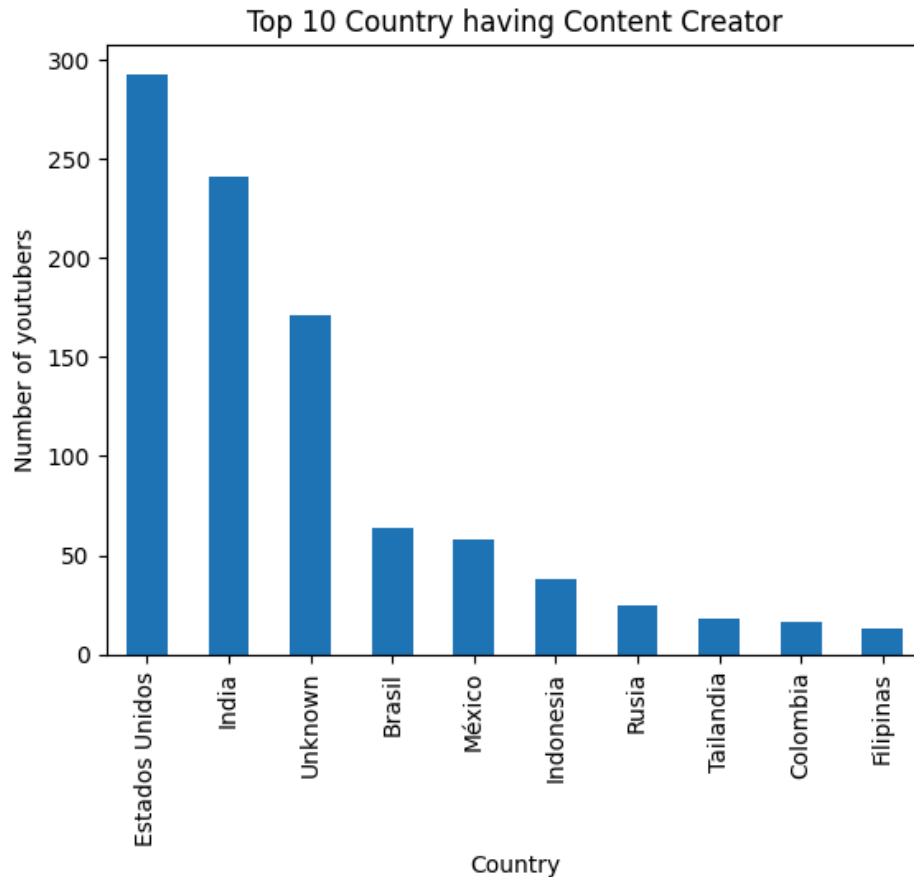
```
>>> 0    Estados Unidos
Name: Country, dtype: object
```

```
youtubers_count=yt['Country'].value_counts().head(10)
youtubers_count
```

```
>>> Country
Estados Unidos    293
India              241
Unknown           171
Brasil            64
México            58
Indonesia         38
Rusia             25
Tailandia         18
Colombia          16
```

```
Filipinas      13
Name: count, dtype: int64
```

```
import matplotlib.pyplot as plt
youtubers_count.plot(kind='bar')
plt.xlabel('Country')
plt.ylabel('Number of youtubers')
plt.title('Top 10 Country having Content Creator')
plt.show()
```



```
Total_Categories=yt['Categories'].value_counts().head(15)
Total_Categories
```



```
Categories
Música y baile      160
Películas, Animación  61
Música y baile, Películas  41
Vlogs diarios      37
Noticias y Política  36
Películas, Humor    34
Animación, Videojuegos  34
Animación, Juguetes  29
Animación, Humor    27
Películas           24
Educación           24
Animación           22
Videojuegos         19
Videojuegos, Humor  17
Música y baile, Animación  16
Name: count, dtype: int64
```

```
yt.dropna(subset = 'Categories', inplace=True)
yt.describe()
```



	Rank	Suscribers	Visits	Likes	Comments
<b>count</b>	694.000000	6.940000e+02	6.940000e+02	6.940000e+02	694.000000
<b>mean</b>	495.298271	2.241556e+07	1.210730e+06	5.347360e+04	1558.793948
<b>std</b>	289.222212	1.824123e+07	6.038274e+06	2.979711e+05	7967.470234
<b>min</b>	1.000000	1.170000e+07	0.000000e+00	0.000000e+00	0.000000
<b>25%</b>	244.250000	1.380000e+07	3.692500e+04	5.685000e+02	2.000000
<b>50%</b>	492.500000	1.680000e+07	1.587000e+05	3.550000e+03	78.000000
<b>75%</b>	746.750000	2.390000e+07	8.339000e+05	2.377500e+04	499.750000
<b>max</b>	1000.000000	2.495000e+08	1.174000e+08	5.300000e+06	154000.000000



```
yt.info()
```



```
<class 'pandas.core.frame.DataFrame'>
Index: 694 entries, 0 to 999
Data columns (total 9 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Rank             694 non-null    int64
1   Username         694 non-null    object
2   Categories       694 non-null    object
3   Suscribers       694 non-null    float64
4   Country          694 non-null    object
5   Visits           694 non-null    float64
6   Likes            694 non-null    float64
7   Comments         694 non-null    float64
8   Links            694 non-null    object
dtypes: float64(4), int64(1), object(4)
memory usage: 54.2+ KB
```

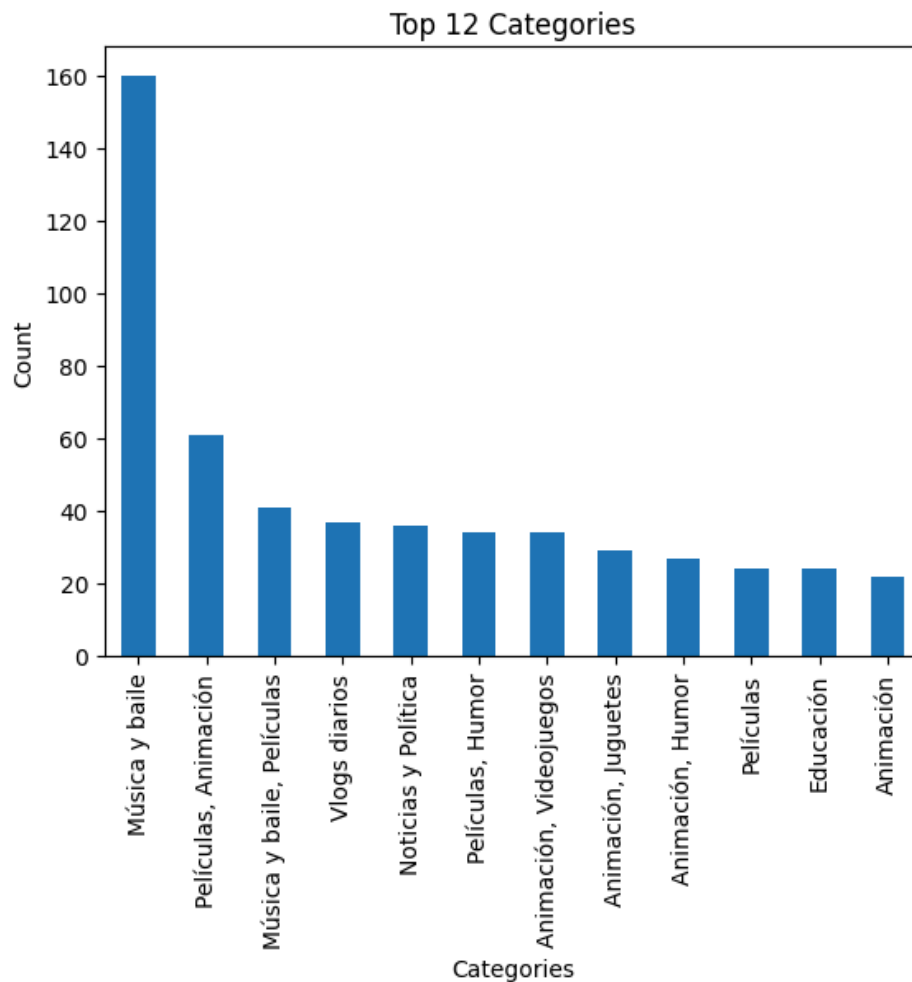
```
Categories=yt['Categories'].unique()
print("Ctegeries in the dataset :")
for Categories in Categories:
    print(Categories)
```



```
Ctegeries in the dataset :
Música y baile
Videojuegos, Humor
Educación
Animación, Juguetes
Películas, Videojuegos
Juguetes
Videojuegos
Películas, Animación
Películas
Noticias y Política
Animación, Humor
Música y baile, Animación
Música y baile, Películas
Películas, Juguetes
Películas, Humor
Vlogs diarios
Videojuegos, Juguetes
Animación, Videojuegos
Animación
Música y baile, Humor
Diseño/arte, DIY y Life Hacks
Ciencia y tecnología
```

Fitness, Salud y autoayuda  
Belleza, Moda  
Humor  
Comida y bebida  
Deportes  
Fitness  
Viajes, Espectáculos  
Comida y bebida, Salud y autoayuda  
Diseño/arte  
DIY y Life Hacks, Juguetes  
Educación, Juguetes  
Juguetes, Coches y vehículos  
Música y baile, Juguetes  
Animales y mascotas  
ASMR  
Moda  
DIY y Life Hacks  
Diseño/arte, Belleza  
Coches y vehículos  
Animación, Humor, Juguetes  
ASMR, Comida y bebida  
Comida y bebida, Juguetes  
Juguetes, DIY y Life Hacks

```
Categories=yt['Categories'].value_counts().head(12)
Categories.plot(kind="bar")
plt.xlabel('Categories')
plt.ylabel("Count")
plt.title("Top 12 Categories")
plt.show()
```



```
suscribers=yt.head(8)
plt.figure(figsize=(10,6))
plt.bar (suscribers['Username'],suscribers['Suscribers'])
plt.xlabel='suscribers'
plt.ylabel='youtubers'
plt.title='Top 8 youtube Suscribers'
plt.show()
```

