

# Importing Libraries

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [2]: #Loading data
df=pd.read_csv("C:\\Users\\vennu\\Downloads\\Task 1 YouTube Streamer Analysis-20240503T080716Z-(1)
df
```

Out[2]:

	Rank	Username	Categories	Suscribers	Country	Visits	Likes	Comments	
0	1	tseries	Música y baile	249500000.0	India	86200.0	2700.0	78.0	http://youtu
1	2	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18500.0	http://youtube.co
2	3	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	0.0	http://youtube
3	4	SETIndia	NaN	162600000.0	India	15600.0	166.0	9.0	http://youtube.c
4	5	KidsDianaShow	Animación, Juguetes	113500000.0	Unknown	3900000.0	12400.0	0.0	http://youtu
...	...	...	...	...	...	...	...	...	
995	996	hamzymukbang	NaN	11700000.0	Estados Unidos	397400.0	14000.0	124.0	http://youtube
996	997	Adaahqueen	NaN	11700000.0	India	1100000.0	92500.0	164.0	http://youtube
997	998	LittleAngelIndonesia	Música y baile	11700000.0	Unknown	211400.0	745.0	0.0	http://youtube
998	999	PenMultiplex	NaN	11700000.0	India	14000.0	81.0	1.0	http://youtube
999	1000	OneindiaHindi	Noticias y Política	11700000.0	India	2200.0	31.0	1.0	http://youtube

1000 rows × 9 columns

```
In [3]: df.set_index('Rank',inplace=True)
df.head()
```

Out[3]:

	Rank	Username	Categories	Suscribers	Country	Visits	Likes	Comments	
	1	tseries	Música y baile	249500000.0	India	86200.0	2700.0	78.0	http://youtube.com/cha
	2	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18500.0	http://youtube.com/channe
	3	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	0.0	http://youtube.com/chan
	4	SETIndia	NaN	162600000.0	India	15600.0	166.0	9.0	http://youtube.com/channe
	5	KidsDianaShow	Animación, Juguetes	113500000.0	Unknown	3900000.0	12400.0	0.0	http://youtube.com/cha

## 1.Exploring data

```
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1000 entries, 1 to 1000
Data columns (total 8 columns):
 #   Column        Non-Null Count  Dtype
---  -
 0   Username      1000 non-null   object
 1   Categories    694 non-null    object
 2   Suscribers    1000 non-null   float64
 3   Country       1000 non-null   object
 4   Visits        1000 non-null   float64
 5   Likes        1000 non-null   float64
 6   Comments      1000 non-null   float64
 7   Links         1000 non-null   object
dtypes: float64(4), object(4)
memory usage: 70.3+ KB
```

```
In [6]: df.columns
```

```
Out[6]: Index(['Username', 'Categories', 'Suscribers', 'Country', 'Visits', 'Likes',
              'Comments', 'Links'],
              dtype='object')
```

```
In [5]: df.dtypes
```

```
Out[5]: Username      object
Categories    object
Suscribers     float64
Country        object
Visits         float64
Likes          float64
Comments       float64
Links          object
dtype: object
```

## Changing datatypes

```
In [8]: df[['Suscribers', 'Visits', 'Likes', 'Comments']] = df[['Suscribers', 'Visits', 'Likes', 'Comments']]
```

```
In [9]: df.dtypes
```

```
Out[9]: Username      object
Categories    object
Suscribers      int32
Country         object
Visits          int32
Likes           int32
Comments        int32
Links           object
dtype: object
```

```
In [6]: df.rename(columns={'Suscribers': 'Subscribers'}, inplace=True)
df.rename(columns={'Username': 'Streamer'}, inplace=True)
```

```
In [7]: df.isnull().sum()
```

```
Out[7]: Streamer      0
Categories    306
Subscribers   0
Country       0
Visits        0
Likes         0
Comments      0
Links         0
dtype: int64
```

## Replacing null values of[Categories] with mode(Most repeated values)

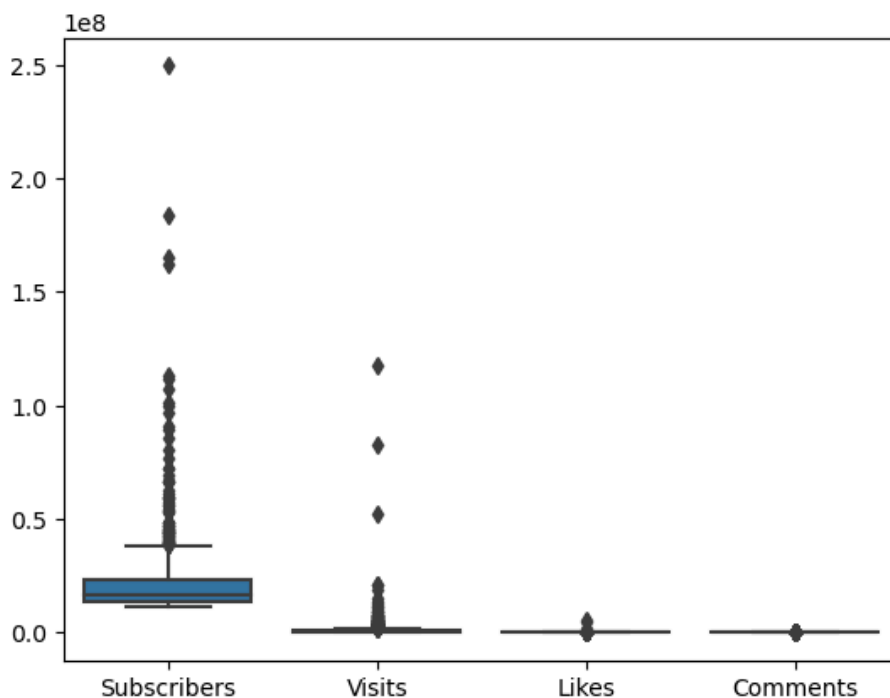
```
In [15]: df['Categories'].fillna(df['Categories'].mode()[0],inplace=True)
```

```
In [16]: df.isnull().sum()
```

```
Out[16]: Streamer      0
Categories    0
Subscribers   0
Country       0
Visits        0
Likes         0
Comments      0
Links         0
dtype: int64
```

```
In [13]: sns.boxplot(data=df)
```

```
Out[13]: <AxesSubplot:>
```




## 2.Trend analysis

```
In [17]: df.head()
```

Out[17]:

	Stremer	Categories	Subscribers	Country	Visits	Likes	Comments	
Rank								
1	tseries	Música y baile	249500000	India	86200	2700	78	http://youtube.com/channe
2	MrBeast	Videojuegos, Humor	183500000	Estados Unidos	117400000	5300000	18500	http://youtube.com/channel/UC
3	CoComelon	Educación	165500000	Unknown	7000000	24700	0	http://youtube.com/channel/I
4	SETIndia	Música y baile	162600000	India	15600	166	9	http://youtube.com/channel/UI
5	KidsDianaShow	Animación, Juguetes	113500000	Unknown	3900000	12400	0	http://youtube.com/channe



## Top 10 Youtube streamers with subscribers

```
In [8]: df.groupby('Streamer')['Subscribers'].sum().sort_values(ascending=False).head(10)
```

Out[8]:

Streamer	
tseries	249500000.0
MrBeast	183500000.0
CoComelon	165500000.0
SETIndia	162600000.0
KidsDianaShow	113500000.0
PewDiePie	111500000.0
LikeNastyaofficial	107500000.0
VladandNiki	101400000.0
zeemusiccompany	99700000.0
WWE	97200000.0

Name: Subscribers, dtype: float64

## Top 10 Youtube Streamers With Views

```
In [9]: df.groupby('Streamer')['Visits'].sum().sort_values(ascending=False).head(10)
```

Out[9]:

Streamer	
MrBeast	117400000.0
MrBeast2	83100000.0
DaFuqBoom	52700000.0
VillageCookingChannel	21500000.0
BeastPhilanthropy	21500000.0
jaanvipatel	19100000.0
_vector_	15400000.0
dojacat	13600000.0
alfredolarin	12900000.0
NickPro	12200000.0

Name: Visits, dtype: float64

## Top 15 Most popular Categories

```
In [11]: df['Categories'].value_counts().sort_values(ascending=False).head(15)
```

```
Out[11]: 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: Categories, dtype: int64
```

## correlation between no of Subscribers,Likes,Comments

```
In [12]: correlation_matrix = df[['Subscribers', 'Likes', 'Comments']].corr()
correlation_matrix
```

```
Out[12]:
```

	Subscribers	Likes	Comments
Subscribers	1.000000	0.211639	0.036350
Likes	0.211639	1.000000	0.325911
Comments	0.036350	0.325911	1.000000

## 3.Distribution of streamers audiences by country,category

```
In [14]: df.groupby(["Country", "Streamer", "Categories"])["Subscribers"].sum().sort_values(ascending=False)
```

```
Out[14]: Country      Streamer      Categories      Subscribers
India      tseries      Música y baile      249500000.0
Estados Unidos  MrBeast      Videojuegos, Humor  183500000.0
Unknown    CoComelon      Educación          165500000.0
           KidsDianaShow  Animación, Juguetes  113500000.0
Estados Unidos  PewDiePie      Películas, Videojuegos  111500000.0
...
Colombia      MykeTowers      Música y baile      11700000.0
Estados Unidos  NFL      Deportes          11700000.0
India      OneindiaHindi      Noticias y Política  11700000.0
Unknown    LittleAngelIndonesia  Música y baile      11700000.0
Estados Unidos  BeAmazed      Educación          11700000.0
Name: Subscribers, Length: 689, dtype: float64
```

```
In [15]: df.head()
```

```
Out[15]:
```

	Streamer	Categories	Subscribers	Country	Visits	Likes	Comments	
Rank								
1	tseries	Música y baile	249500000.0	India	86200.0	2700.0	78.0	http://youtube.com/cha
2	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18500.0	http://youtube.com/channe
3	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	0.0	http://youtube.com/chan
4	SETIndia	NaN	162600000.0	India	15600.0	166.0	9.0	http://youtube.com/channe
5	KidsDianaShow	Animación, Juguetes	113500000.0	Unknown	3900000.0	12400.0	0.0	http://youtube.com/cha

## 4. Performance metrics

```
In [16]: Avg_subs=df['Subscribers'].mean()  
Avg_subs
```

```
Out[16]: 21894400.0
```

```
In [47]: df['Visits'].mean()
```

```
Out[47]: 1209446.3155
```

```
In [18]: df['Likes'].mean()
```

```
Out[18]: 53632.592
```

```
In [19]: df['Comments'].mean()
```

```
Out[19]: 1288.768
```

## 5. Content Categories

```
In [20]: df.groupby('Categories')['Streamer'].count().sort_values(ascending=False)
```

```
Out[20]: Categories
Música y baile                160
Películas, Animación          61
Música y baile, Películas     41
Vlogs diarios                 37
Noticias y Política           36
Animación, Videojuegos        34
Películas, Humor              34
Animación, Juguetes           29
Animación, Humor              27
Educación                     24
Películas                     24
Animación                     22
Videojuegos                   19
Videojuegos, Humor            17
Música y baile, Animación      16
Ciencia y tecnología           14
Comida y bebida                12
Juguetes                       10
Humor                         10
Películas, Juguetes            9
Deportes                       8
Películas, Videojuegos         8
Música y baile, Humor          6
Juguetes, Coches y vehículos   4
Videojuegos, Juguetes          3
Fitness, Salud y autoayuda      3
DIY y Life Hacks               3
Moda                           2
Fitness                        2
Educación, Juguetes            2
Coches y vehículos            2
Animales y mascotas            2
Juguetes, DIY y Life Hacks      1
ASMR, Comida y bebida          1
Música y baile, Juguetes       1
Diseño/arte, DIY y Life Hacks   1
Diseño/arte, Belleza           1
Diseño/arte                    1
DIY y Life Hacks, Juguetes      1
Comida y bebida, Salud y autoayuda 1
Comida y bebida, Juguetes       1
Viajes, Espectáculos           1
Belleza, Moda                  1
Animación, Humor, Juguetes      1
ASMR                           1
Name: Streamer, dtype: int64
```

## 6.Brands and collaborations

While high subscriber counts can certainly attract brands, other factors like engagement rate, audience demographics, and content alignment with brand values also play crucial roles in determining collaboration opportunities.

```
-->T-Series: 249.5 million
MrBeast: 183.5 million
CoComelon: 16.5 million
SET India: 162.6 million
Kids Diana Show:113.5 million
PewDiePi: 111.5 million
Like Nastya Official: 107.5 million
Vlad and Niki: 101.4 million
Zee Music Company:99.7 million
WWE: 97.2 million
```

With these Youtube Streamers we can promote brands, collaboration with related content field, and also we can do market compaigns.

## 7.Streamers with above average Performance

```
In [21]: Most_avg_values =df[df['Subscribers'] > df['Subscribers'].mean()][ 'Subscribers']
Most_avg_values
```

```
Out[21]: Rank
1      249500000.0
2      183500000.0
3      165500000.0
4      162600000.0
5      113500000.0
...
299    21900000.0
300    21900000.0
301    21900000.0
302    21900000.0
303    21900000.0
Name: Subscribers, Length: 303, dtype: float64
```

```
In [29]: avg_subscribers_by_Streamer = df.groupby('Streamer')['Subscribers'].transform('mean')
avg_subscribers_by_Streamer
```

```
Out[29]: Rank
1      249500000.0
2      183500000.0
3      165500000.0
4      162600000.0
5      113500000.0
...
996    11700000.0
997    11700000.0
998    11700000.0
999    11700000.0
1000   11700000.0
Name: Subscribers, Length: 1000, dtype: float64
```

```
In [30]: df['Subs Greater than avg']= df['Subscribers'] > avg_subscribers_by_Streamer
```

```
In [32]: df_subs_above_avg = df[df['Subs Greater than avg'] == True]
df_subs_above_avg
```

Out[32]:

	Streamer	Categories	Subscribers	Country	Visits	Likes	Comments
Rank							
451	thexoteam	NaN	17900000.0	Estados Unidos	772800.0	45000.0	185.0 <a href="http://youtube.com/channel/UCIZAOlfhJt">http://youtube.com/channel/UCIZAOlfhJt</a>



```
In [33]: Most_avg_likes = df[df['Likes'] > df['Likes'].mean()]['Likes']
Most_avg_likes

avg_likes_by_streamer = df.groupby('Streamer')['Likes'].transform('mean')

df['Likes Greater than avg'] = df['Likes'] > avg_likes_by_streamer

df_likes_above_avg = df[df['Likes Greater than avg'] == True]
df_likes_above_avg
```

Out[33]:

	Streamer	Categories	Subscribers	Country	Visits	Likes	Comments	
Rank								
448	mgcplayhouse	Juguetes, Coches y vehículos	17800000.0	Unknown	56300.0	96.0	0.0	<a href="http://youtube.com/channel/UC6zF">http://youtube.com/channel/UC6zF</a>
450	thexoteam	NaN	17800000.0	Estados Unidos	797600.0	50400.0	179.0	<a href="http://youtube.com/channel/UCIZAC">http://youtube.com/channel/UCIZAC</a>
950	Family-Box	Películas	12000000.0	Rusia	173600.0	6600.0	105.0	<a href="http://youtuhjHN">http://youtuhjHN</a>

```
In [34]: Most_avg_visits = df[df['Visits'] > df['Visits'].mean()]['Visits']
Most_avg_visits

avg_visits_by_streamer = df.groupby('Streamer')['Visits'].transform('mean')

df['Visits Greater than avg'] = df['Visits'] > avg_visits_by_streamer

df_visits_above_avg = df[df['Visits Greater than avg'] == True]
df_visits_above_avg
```

Out[34]:

	Streamer	Categories	Subscribers	Country	Visits	Likes	Comments	
Rank								
450	thexoteam	NaN	17800000.0	Estados Unidos	797600.0	50400.0	179.0	<a href="http://youtube.com/channel/UCIZAC">http://youtube.com/channel/UCIZAC</a>
452	mgcplayhouse	Juguetes, Coches y vehículos	17800000.0	Unknown	63600.0	75.0	0.0	<a href="http://youtube.com/channel/UC6zF">http://youtube.com/channel/UC6zF</a>
957	Family-Box	Películas	12000000.0	Rusia	177400.0	6300.0	86.0	<a href="http://youtuhjHN">http://youtuhjHN</a>

```
In [35]: Most_avg_comments = df[df['Comments'] > df['Comments'].mean()]['Comments']
Most_avg_comments

avg_comments_by_streamer = df.groupby('Streamer')['Comments'].transform('mean')

df['Comments Greater than avg'] = df['Comments'] > avg_comments_by_streamer

df_comments_above_avg = df[df['Comments Greater than avg'] == True]
df_comments_above_avg
```

Out[35]:

	Streamer	Categories	Subscribers	Country	Visits	Likes	Comments
k							
1	thexoteam	NaN	17900000.0	Estados Unidos	772800.0	45000.0	185.0 <a href="http://youtube.com/channel/UCIZAOfhJQJRym">http://youtube.com/channel/UCIZAOfhJQJRym</a>
0	Family-Box	Películas	12000000.0	Rusia	173600.0	6600.0	105.0 <a href="http://youtube.com/channel/UCjHNWViReG6R">http://youtube.com/channel/UCjHNWViReG6R</a>

```
In [36]: columns_to_replace = ['Subs Greater than avg', 'Likes Greater than avg', 'Visits Greater than a

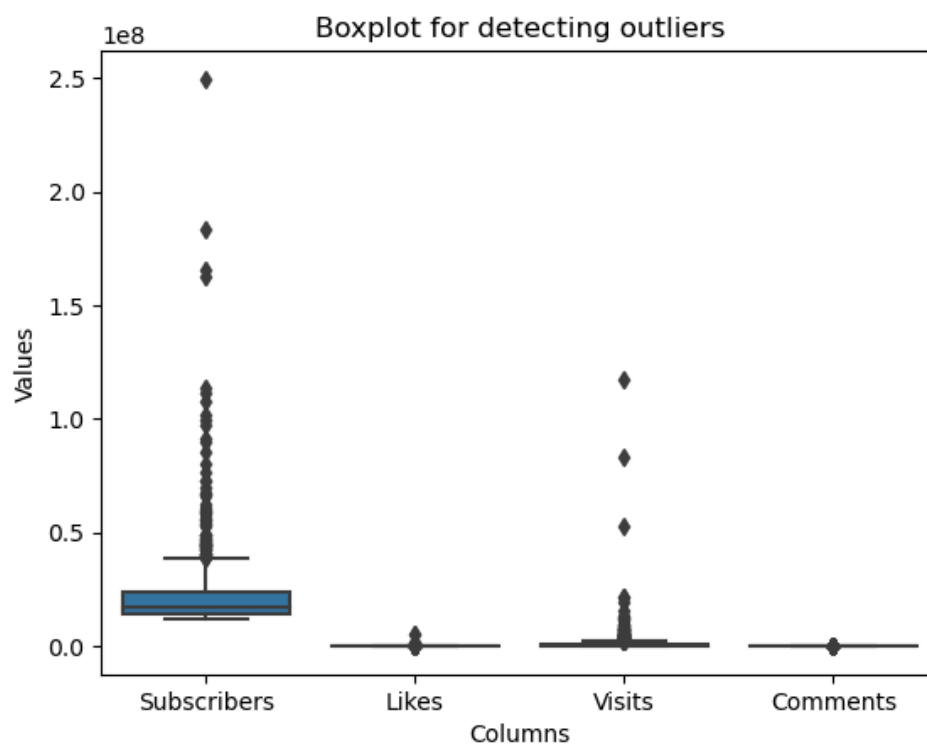
for column in columns_to_replace:
    df[column] = df[column].replace({True:1, False: 0})
```

```
In [37]: import seaborn as sns
import matplotlib.pyplot as plt

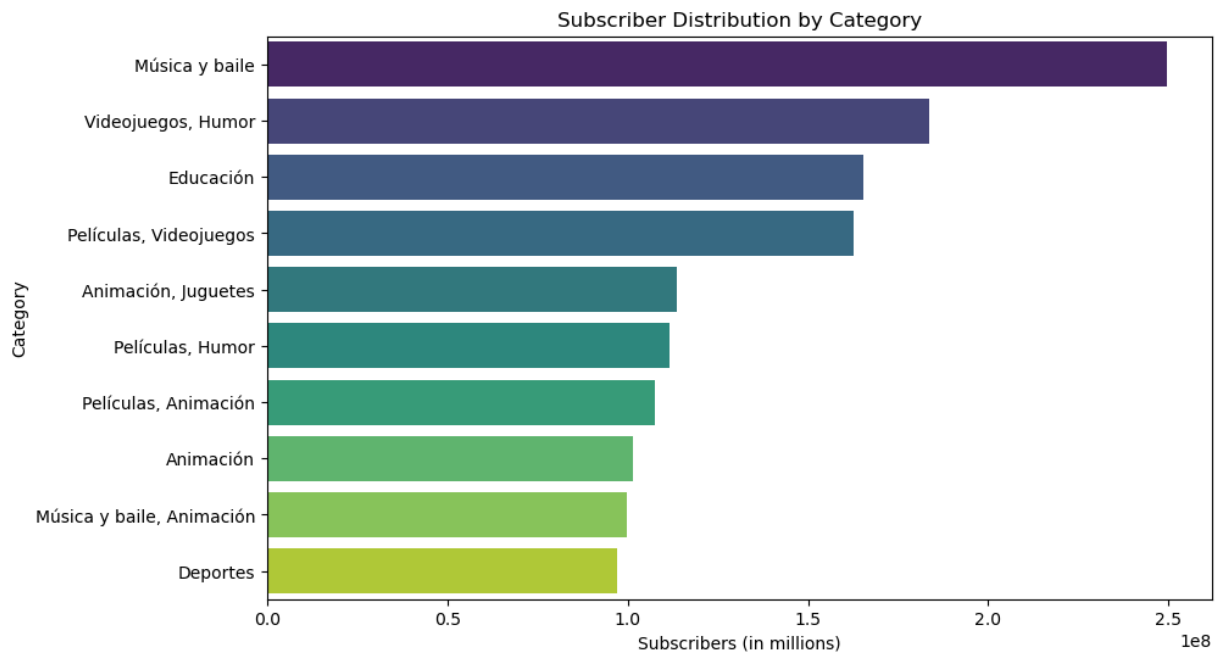
columns_to_plot = ["Subscribers", "Likes", "Visits", "Comments"]

sns.boxplot(data=df[columns_to_plot])

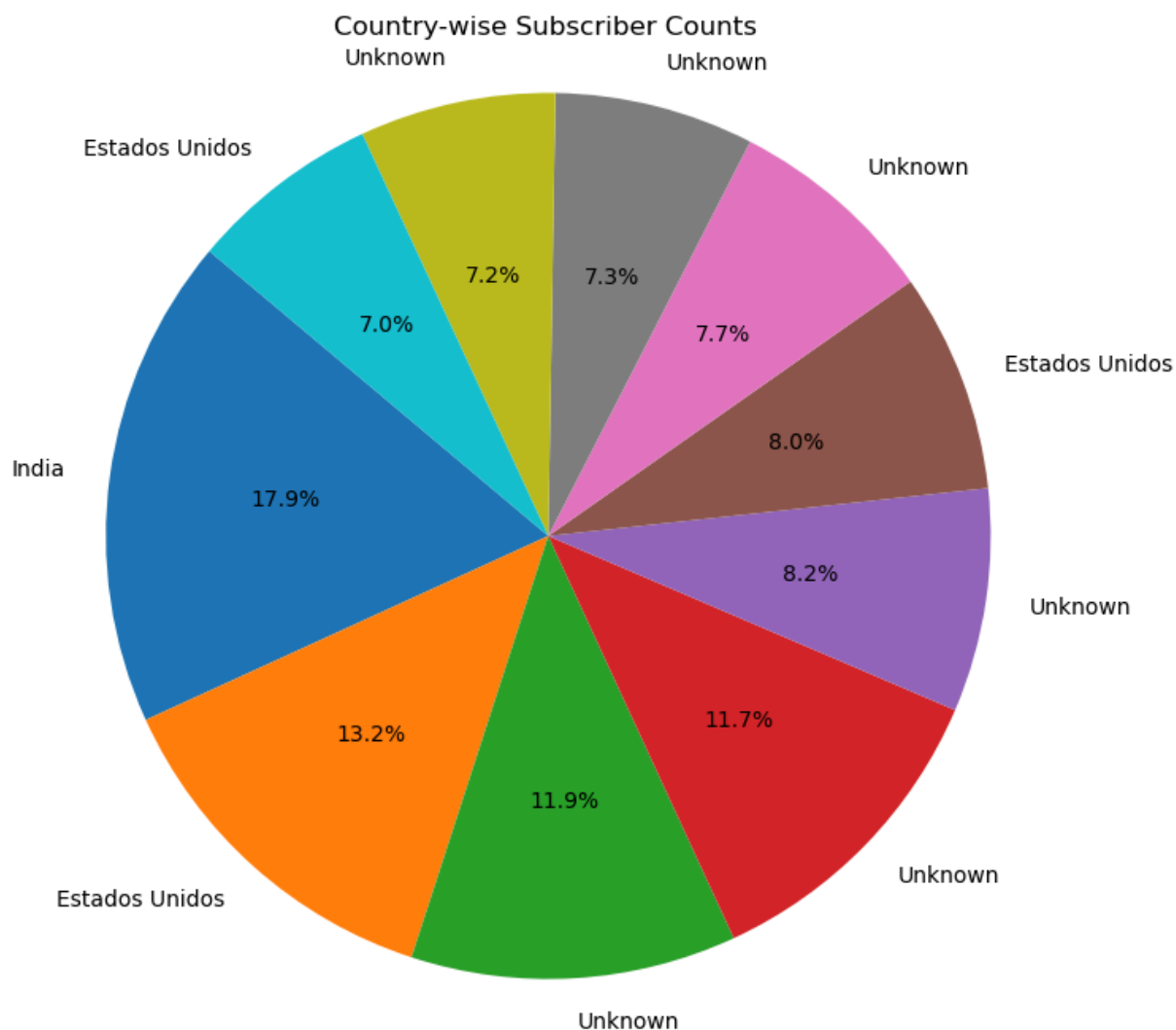
plt.title('Boxplot for detecting outliers')
plt.xlabel('Columns')
plt.ylabel('Values')
plt.show()
```



```
In [38]: streamers = ['T-Series', 'MrBeast', 'CoComelon', 'SETIndia', 'KidsDianaShow', 'PewDiePie', 'Lik
categories = ['Música y baile', 'Videojuegos, Humor', 'Educación', 'Películas, Videojuegos', 'A
subscribers = [249500000, 183500000, 165500000, 162600000, 113500000, 111500000, 107500000, 101
country = ['India', 'Estados Unidos', 'Unknown', 'Unknown', 'Unknown', 'Estados Unidos', 'Unkno
plt.figure(figsize=(10, 6))
sns.barplot(x=subscribers, y=categories, palette='viridis')
plt.title('Subscriber Distribution by Category')
plt.xlabel('Subscribers (in millions)')
plt.ylabel('Category')
plt.show()
```



```
In [39]: plt.figure(figsize=(8, 8))
plt.pie(subscribers, labels=country, autopct='%1.1f%%', startangle=140)
plt.title('Country-wise Subscriber Counts')
plt.axis('equal')
plt.show()
```



## 8. Insights and recommendation

-->Here are Top 10 Youtube Streamers with their subscribers:

T-Series: 249.5 million  
 MrBeast: 183.5 million  
 CoComelon: 165.5 million  
 SET India: 162.6 million  
 Kids Diana Show: 113.5 million  
 PewDiePie: 111.5 million  
 Like Nastya Official: 107.5 million  
 Vlad and Niki: 101.4 million  
 Zee Music Company: 99.7 million  
 WWE: 97.2 million

## Top 10 Youtube channels with their link

```
In [41]: df.groupby(['Streamer', 'Links'])['Subscribers'].sum().sort_values(ascending=False).head(10)
```

```
Out[41]: Streamer      Links
tseries      http://youtube.com/channel/UCq-Fj5jknLsUf-MWSy4_brA (http://youtube.com/ch
annel/UCq-Fj5jknLsUf-MWSy4_brA) 249500000.0
MrBeast      http://youtube.com/channel/UCX60Q3DkcsbYNE6H8uQQuVA (http://youtube.com/ch
annel/UCX60Q3DkcsbYNE6H8uQQuVA) 183500000.0
CoComelon    http://youtube.com/channel/UCbCmjCuTUZos6Inko4u57UQ (http://youtube.com/ch
annel/UCbCmjCuTUZos6Inko4u57UQ) 165500000.0
SETIndia     http://youtube.com/channel/UCpEhnqL0y41EpW2TvWAHD7Q (http://youtube.com/ch
annel/UCpEhnqL0y41EpW2TvWAHD7Q) 162600000.0
KidsDianaShow http://youtube.com/channel/UCk8GzjM0rta8yxDckfy1JYw (http://youtube.com/ch
annel/UCk8GzjM0rta8yxDckfy1JYw) 113500000.0
PewDiePie    http://youtube.com/channel/UC-lHjZR3Gqxm24_Vd_AJ5Yw (http://youtube.com/ch
annel/UC-lHjZR3Gqxm24_Vd_AJ5Yw) 111500000.0
LikeNastyaoofficial http://youtube.com/channel/UCJp1p5SjeGSdVdwsfb9Q71Q (http://youtube.com/ch
annel/UCJp1p5SjeGSdVdwsfb9Q71Q) 107500000.0
VladandNiki  http://youtube.com/channel/UCv1E5gTb0vjio1F1Em-c_Ow (http://youtube.com/ch
annel/UCv1E5gTb0vjio1F1Em-c_Ow) 101400000.0
zeemusiccompany http://youtube.com/channel/UCFFbwnve3yF62-tVXkTyHqg (http://youtube.com/ch
annel/UCFFbwnve3yF62-tVXkTyHqg) 99700000.0
WWE          http://youtube.com/channel/UCJ5v_MCY6GNUMBT08-D3XoAg (http://youtube.com/ch
annel/UCJ5v_MCY6GNUMBT08-D3XoAg) 97200000.0
Name: Subscribers, dtype: float64
```

-->Here are the Top 15 popular 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

#Regional preferences

-->Indian Preference for Music and Dance Content.  
-->US Influence in Gaming and Entertainment.  
-->Global Appeal of Educational Content:  
CoComelon, a channel focusing on educational content, garners a substantial audience despite its origin being unknown. This suggests that educational content transcends geographical boundaries and has universal appeal.  
-->Entertainment for Children Across Regions:  
KidsDianaShow, representing an unknown origin, gathers a considerable audience with content focused on animation and toys ("Animación, Juguetes"). This indicates a global interest in entertainment content tailored for children.

#Countries with most no of subscribers

-->India:T-Series: Música y baile (249.5 million subscribers)  
-->CoComelon: Educación (165.5 million subscribers)  
-->United States:MrBeast: Videojuegos, Humor (183.5 million subscribers)

#Average number of subscribers, visits, likes, and comments.

Average Subscribers: 21.89 million  
Average Visits: 1.21 million  
Average Likes: 53.63 thousand  
Average Comments: 1.29 thousand.

#### Categories with the Highest Number of Streamers:

Música y baile	160
Películas, Animación	61
Música y baile, Películas	41
Vlogs diarios	37
Noticias y Política	36
Animación, Videojuegos	34
Películas, Humor	34
Animación, Juguetes	29
Animación, Humor	27
Educación	24
Películas	24
Animación	22
Videojuegos	19
Videojuegos, Humor	17
Música y baile, Animación	16
Ciencia y tecnología	14

#### Benchmarking performers

->Out of 1000 streamers, 1 streamer has above the average subscriber count.  
-->Three streamers, namely MGCPlayhouse, TheXOteam, and Family-Box, have above-average likes, visits, and comments.  
-->These streamers not only possess a significant subscriber base but also enjoy higher engagement levels compared to the average, making them potentially valuable collaborators for brand partnerships and marketing campaigns.

#### Outlier values

presence of outliers in subscriber counts and visit numbers suggests that certain highly popular channels are experiencing exceptionally high levels of engagement and viewership. These outliers indicate that a select few channels are attracting an unusually large number of subscribers and visits compared to the majority of channels.

**\*\*While high subscriber counts can certainly attract brands, other factors like engagement rate, audience demographics, and content alignment with brand values also play crucial roles in determining collaboration opportunities.**

-->T-Series: 249.5 million

MrBeast: 183.5 million

CoComelon: 165.5 million

SET India: 162.6 million

Kids Diana Show: 113.5 million

PewDiePie: 111.5 million

Like Nastya Official: 107.5 million

Vlad and Niki: 101.4 million

Zee Music Company: 99.7 million

WWE: 97.2 million

With these youtube streamers we can promote brands, collaboration with related content field, and also we can do market campaigns.