Importing Libraries

In [2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
C:\Users\91778\anaconda3\lib\site-packages\scipy\__init__.py:155: UserWarn
ing: A NumPy version >=1.18.5 and <1.25.0 is required for this version of
SciPy (detected version 1.26.2
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}"
```

In [5]:
```
# Loading data
df=pd.read_csv("C:\\Users\\91778\\Downloads\\youtubers_df.csv")
df
```

Out[5]:

	Rank	Username	Categories	Suscribers	Country	Visits	Likes
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
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
...	...	...	...	...	...	...	...
995	996	hamzymukbang	NaN	11700000.0	Estados Unidos	397400.0	14000.0
996	997	Adaahqueen	NaN	11700000.0	India	1100000.0	92500.0
997	998	LittleAngelIndonesia	Música y baile	11700000.0	Unknown	211400.0	745.0
998	999	PenMultiplex	NaN	11700000.0	India	14000.0	81.0
999	1000	OneindiaHindi	Noticias y Política	11700000.0	India	2200.0	31.0

1000 rows × 9 columns

In [6]:
```
df.set_index('Rank',inplace=True)
df.head()
```

Out[6]:

Rank	Username	Categories	Suscribers	Country	Visits	Likes	Comments
1	tseries	Música y baile	249500000.0	India	86200.0	2700.0	78.(
2	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18500.(
3	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	0.(
4	SETIndia	NaN	162600000.0	India	15600.0	166.0	9.(
5	KidsDianaShow	Animación, Juguetes	113500000.0	Unknown	3900000.0	12400.0	0.(

1) Exploring data

In [7]:
```
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 [79]:
```
df.columns
```

Out[79]:
```
Index(['Username', 'Categories', 'Suscribers', 'Country', 'Visits', 'Like
s',
       'Comments', 'Links'],
      dtype='object')
```

In [80]:
```
df.dtypes
```

Out[80]:
```
Username       object
Categories     object
Suscribers    float64
Country        object
Visits        float64
Likes         float64
Comments      float64
Links          object
dtype: object
```

Changing datatypes

In [81]:
```
df[['Suscribers', 'Visits', 'Likes', 'Comments']] = df[['Suscribers', 'Visi
```

In [82]:
```
df.dtypes
```

Out[82]:
```
Username      object
Categories    object
Suscribers     int32
Country       object
Visits         int32
Likes          int32
Comments       int32
Links         object
dtype: object
```

In [83]:
```
df.rename(columns={'Suscribers': 'Subscribers'}, inplace=True)
df.rename(columns={'Username' : 'Streamer'},inplace=True)
```

In [84]:
```
df.isnull().sum()
```

Out[84]:
```
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 [85]:
```
df['Categories'].fillna(df['Categories'].mode()[0],inplace=True)
```

In [86]:
```
df.isnull().sum()
```

Out[86]:
```
Streamer       0
Categories     0
Subscribers    0
Country        0
Visits         0
Likes          0
Comments       0
Links          0
dtype: int64
```

2) Trend analysis

In [87]:
```
df.head()
```

Out[87]:

Rank	Streamer	Categories	Subscribers	Country	Visits	Likes	Comments
1	tseries	Música y baile	249500000	India	86200	2700	78
2	MrBeast	Videojuegos, Humor	183500000	Estados Unidos	117400000	5300000	18500
3	CoComelon	Educación	165500000	Unknown	7000000	24700	0
4	SETIndia	Música y baile	162600000	India	15600	166	9
5	KidsDianaShow	Animación, Juguetes	113500000	Unknown	3900000	12400	0

Top10 youtube streamers with subscribers

In [88]:
```
df.groupby('Streamer')['Subscribers'].sum().sort_values(ascending=False).he
```

Out[88]:
```
Streamer
tseries                249500000
MrBeast                183500000
CoComelon              165500000
SETIndia               162600000
KidsDianaShow          113500000
PewDiePie              111500000
LikeNastyaofficial     107500000
VladandNiki            101400000
zeemusiccompany         99700000
WWE                     97200000
Name: Subscribers, dtype: int32
```

Top10 youtube streamers with Views

In [89]:
```
df.groupby('Streamer')['Visits'].sum().sort_values(ascending=False).head(10
```

Out[89]:
```
Streamer
MrBeast                  117400000
MrBeast2                  83100000
DaFuqBoom                 52700000
VillageCookingChannel     21500000
BeastPhilanthropy         21500000
jaanvipatel               19100000
_vector_                  15400000
dojacat                   13600000
alfredolarin              12900000
NickPro                   12200000
Name: Visits, dtype: int32
```

Top15 Most popular categories

In [90]:
```
df['Categories'].value_counts().sort_values(ascending=False).head(15)
```

Out[90]:
```
Música y baile                466
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 [91]:
```
correlation_matrix = df[['Subscribers', 'Likes', 'Comments']].corr()
correlation_matrix
```

Out[91]:

	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

In [92]:
```
df.groupby(["Country","Streamer","Categories"])["Subscribers"].sum().sort_v
```

Out[92]:
```
Country         Streamer       Categories
India           tseries        Música y baile         249500000
Estados Unidos  MrBeast        Videojuegos, Humor     183500000
Unknown         CoComelon      Educación              165500000
India           SETIndia       Música y baile         162600000
Unknown         KidsDianaShow  Animación, Juguetes    113500000
                                                          ...
Estados Unidos  cut            Humor                   11700000
                hamzymukbang   Música y baile          11700000
India           VYRLOriginals  Música y baile          11700000
Estados Unidos  BeAmazed       Educación               11700000
Colombia        MykeTowers     Música y baile          11700000
Name: Subscribers, Length: 994, dtype: int32
```

In [93]:
```
df.head()
```

Out[93]:

Rank	Streamer	Categories	Subscribers	Country	Visits	Likes	Comments
1	tseries	Música y baile	249500000	India	86200	2700	78
2	MrBeast	Videojuegos, Humor	183500000	Estados Unidos	117400000	5300000	18500
3	CoComelon	Educación	165500000	Unknown	7000000	24700	0
4	SETIndia	Música y baile	162600000	India	15600	166	9
5	KidsDianaShow	Animación, Juguetes	113500000	Unknown	3900000	12400	0

4) Performance metrics

In [94]:
```
Avg_subs=df['Subscribers'].mean()
Avg_subs
```

Out[94]:
```
21894399.987
```

In [95]:
```
df['Visits'].mean()
```

Out[95]:
```
1209446.315
```

In [96]:
```
df['Likes'].mean()
```

Out[96]:
```
53632.592
```

In [97]:
```
df['Comments'].mean()
```

Out[97]:
```
1288.768
```

5) Content categories

In [98]:
```
df.groupby('Categories')["Streamer"].count().sort_values(ascending=False)
```

Out[98]:
```
Categories
Música y baile                          466
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: 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.
```

7) Streamers with above average performance

In [99]:
```
Most_avg_values =df[df['Subscribers'] > df['Subscribers'].mean()]['Subscrib
Most_avg_values
```

Out[99]:
```
Rank
1      249500000
2      183500000
3      165500000
4      162600000
5      113500000
         ...
299     21900000
300     21900000
301     21900000
302     21900000
303     21900000
Name: Subscribers, Length: 303, dtype: int32
```

In [100]:
```
avg_subscribers_by_Streamer = df.groupby('Streamer')['Subscribers'].transfo
avg_subscribers_by_Streamer
```

Out[100]:
```
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 [101]:
```
df['Subs Greater than avg'] = df['Subscribers'] > avg_subscribers_by_Stream
```

In [102]:
```
df_subs_above_avg = df[df['Subs Greater than avg'] == True]
df_subs_above_avg
```

Out[102]:

Rank	Streamer	Categories	Subscribers	Country	Visits	Likes	Comments	
451	thexoteam	Música y baile	17900000	Estados Unidos	772800	45000	185	http://youtube.

In [103]:
```
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[103]:

Rank	Streamer	Categories	Subscribers	Country	Visits	Likes	Comments	
448	mgcplayhouse	Juguetes, Coches y vehículos	17800000	Unknown	56300	96	0	http://you
450	thexoteam	Música y baile	17800000	Estados Unidos	797600	50400	179	http://you
950	Family-Box	Películas	12000000	Rusia	173600	6600	105	

In [104]:
```
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[104]:

Rank	Streamer	Categories	Subscribers	Country	Visits	Likes	Comments	
450	thexoteam	Música y baile	17800000	Estados Unidos	797600	50400	179	http://you
452	mgcplayhouse	Juguetes, Coches y vehículos	17800000	Unknown	63600	75	0	http://you
957	Family-Box	Películas	12000000	Rusia	177400	6300	86	

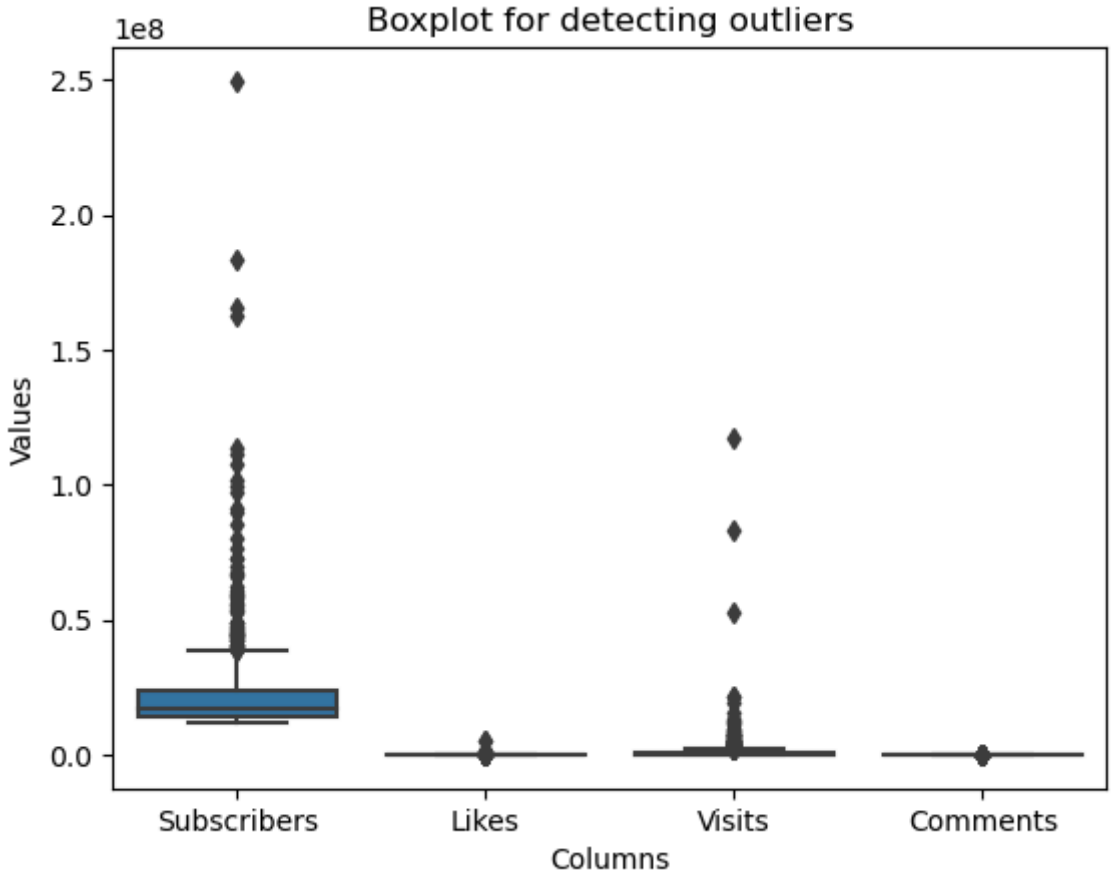
In [105]:
```
Most_avg_comments = df[df['Comments'] > df['Comments'].mean()]['Comments']
Most_avg_comments

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

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[105]:

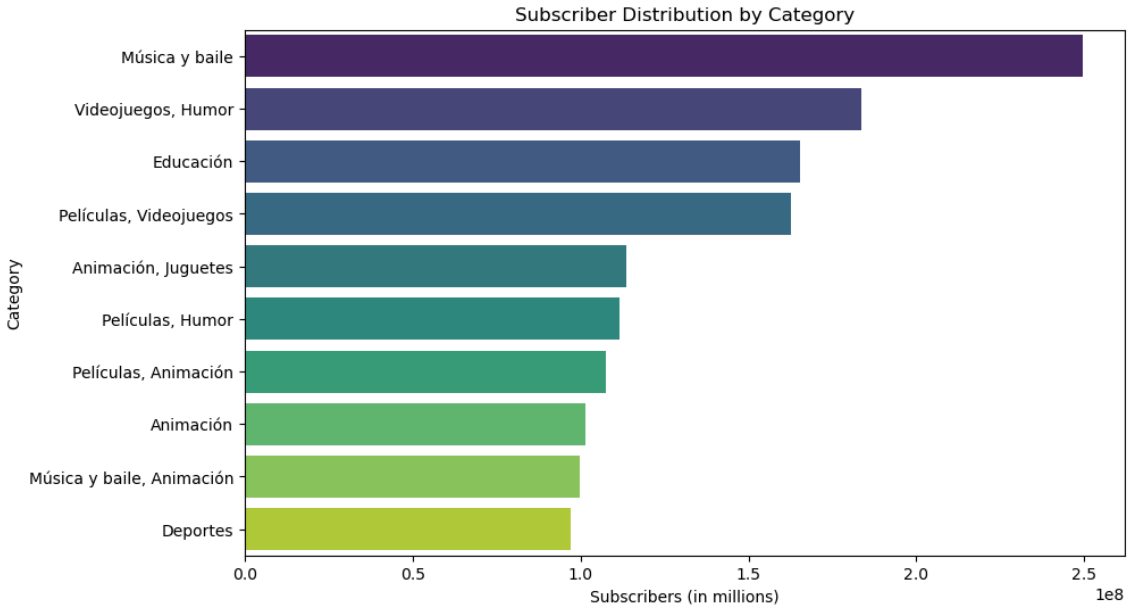
Rank	Streamer	Categories	Subscribers	Country	Visits	Likes	Comments	
451	thexoteam	Música y baile	17900000	Estados Unidos	772800	45000	185	http://youtube.
950	Family-Box	Películas	12000000	Rusia	173600	6600	105	

In [106]:
```
columns_to_replace = ['Subs Greater than avg', 'Likes Greater than avg', 'V

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

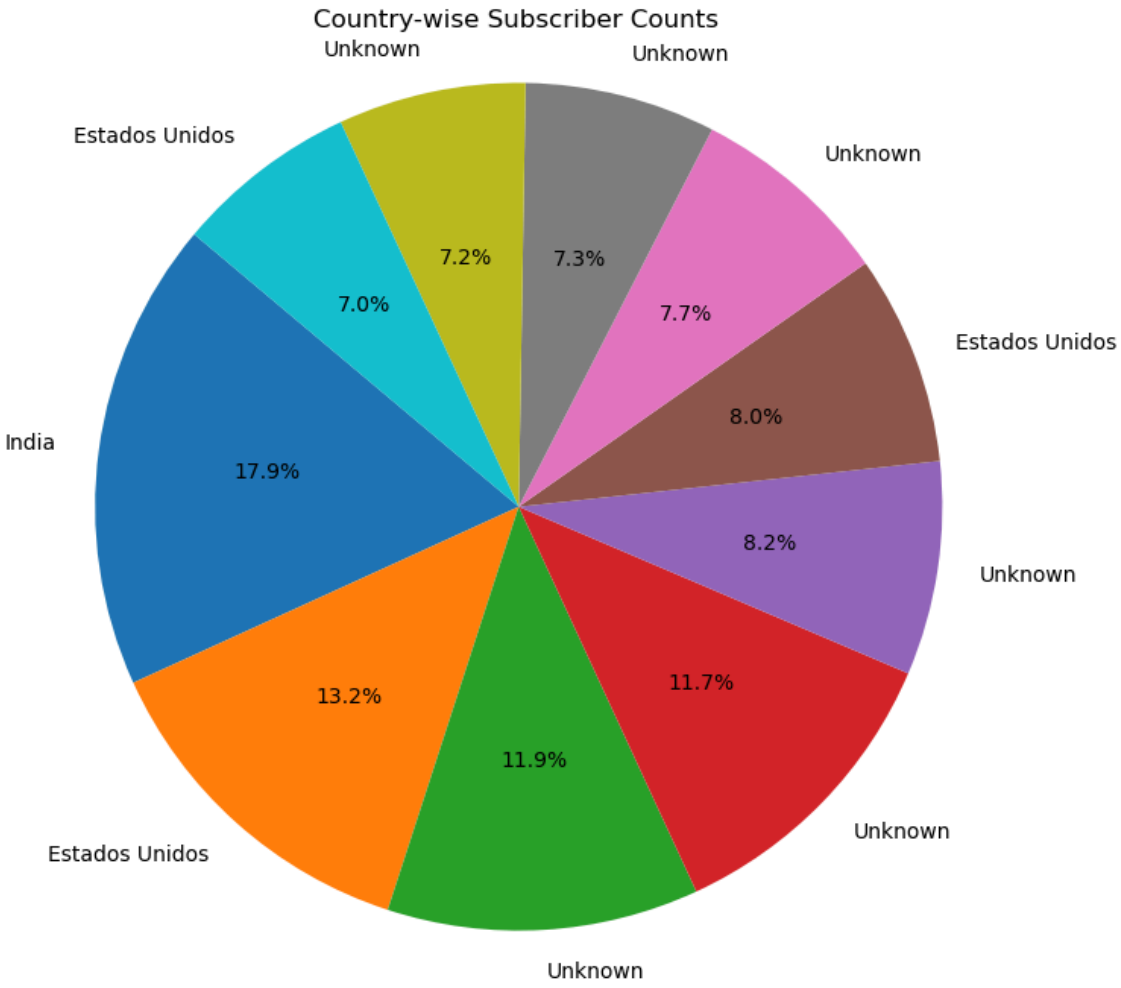
In [107]:
```
import seaborn as sns
import matplotlib.pyplot as plt

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()
```



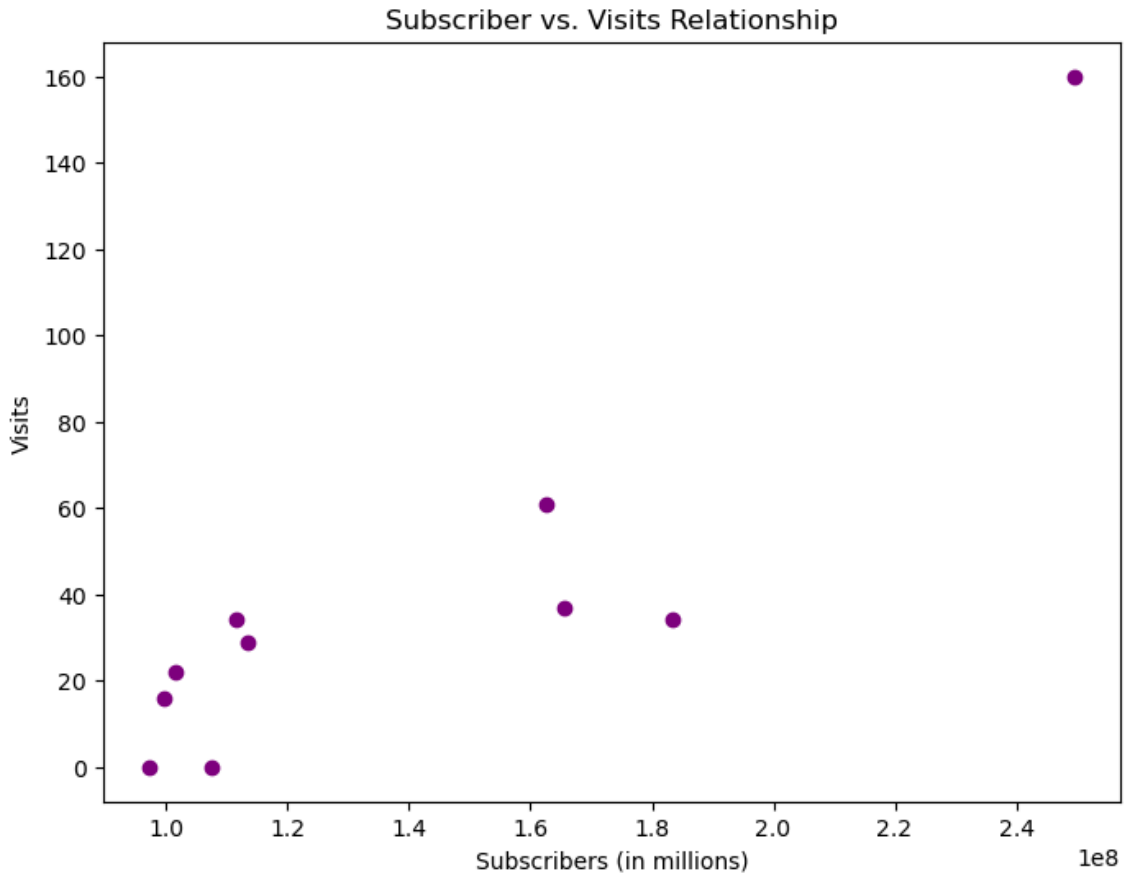
In [108]:
```
streamers = ['T-Series', 'MrBeast', 'CoComelon', 'SETIndia', 'KidsDianaShow
categories = ['Música y baile', 'Videojuegos, Humor', 'Educación', 'Películ
subscribers = [249500000, 183500000, 165500000, 162600000, 113500000, 11150
country = ['India', 'Estados Unidos', 'Unknown', 'Unknown', 'Unknown', 'Est
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 [109]:
```
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()
```



In [110]:
```
# Plot 4: Subscriber vs. Visits Relationship
plt.figure(figsize=(8, 6))
plt.scatter(subscribers, visits, color='purple')
plt.title('Subscriber vs. Visits Relationship')
plt.xlabel('Subscribers (in millions)')
plt.ylabel('Visits')
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 [116]:
```
df.groupby(['Streamer','Links'])['Subscribers'].sum().sort_values(ascending
```

Out[116]:
```
Streamer            Links
tseries             http://youtube.com/channel/UCq-Fj5jknLsUf-MWSy4_brA (h
ttp://youtube.com/channel/UCq-Fj5jknLsUf-MWSy4_brA)     249500000
MrBeast             http://youtube.com/channel/UCX6OQ3DkcsbYNE6H8uQQuVA (h
ttp://youtube.com/channel/UCX6OQ3DkcsbYNE6H8uQQuVA)     183500000
CoComelon           http://youtube.com/channel/UCbCmjCuTUZos6Inko4u57UQ (h
ttp://youtube.com/channel/UCbCmjCuTUZos6Inko4u57UQ)     165500000
SETIndia            http://youtube.com/channel/UCpEhnqL0y41EpW2TvWAHD7Q (h
ttp://youtube.com/channel/UCpEhnqL0y41EpW2TvWAHD7Q)     162600000
KidsDianaShow       http://youtube.com/channel/UCk8GzjMOrta8yxDcKfylJYw (h
ttp://youtube.com/channel/UCk8GzjMOrta8yxDcKfylJYw)     113500000
PewDiePie           http://youtube.com/channel/UC-lHJZR3Gqxm24_Vd_AJ5Yw (h
ttp://youtube.com/channel/UC-lHJZR3Gqxm24_Vd_AJ5Yw)     111500000
LikeNastyaofficial  http://youtube.com/channel/UCJplp5SjeGSdVdwsfb9Q7lQ (h
ttp://youtube.com/channel/UCJplp5SjeGSdVdwsfb9Q7lQ)     107500000
VladandNiki         http://youtube.com/channel/UCvlE5gTbOvjiolFlEm-c_Ow (h
ttp://youtube.com/channel/UCvlE5gTbOvjiolFlEm-c_Ow)     101400000
zeemusiccompany     http://youtube.com/channel/UCFFbwnve3yF62-tVXkTyHqg (h
ttp://youtube.com/channel/UCFFbwnve3yF62-tVXkTyHqg)      99700000
WWE                 http://youtube.com/channel/UCJ5v_MCY6GNUBTO8-D3XoAg (h
ttp://youtube.com/channel/UCJ5v_MCY6GNUBTO8-D3XoAg)      97200000
Name: Subscribers, dtype: int32
```

-->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

```
The 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.

In [ ]:

In [ ]: