

Exploratory Data Analysis (EDA) of YouTube and Spotify Datasets

Libraries import

```
In [3]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px

In [4]: df = pd.read_csv(r"C:\Users\Dell\Downloads\you tube and spotify analysis\Spotify Youtube Dataset.csv")

In [5]: df.columns

Out[5]: Index(['Unnamed: 0', 'Artist', 'Url_spotify', 'Track', 'Album', 'Album_type',
              'Uri', 'Danceability', 'Energy', 'Key', 'Loudness', 'Speechiness',
              'Acousticness', 'Instrumentalness', 'Liveness', 'Valence', 'Tempo',
              'Duration_ms', 'Url_youtube', 'Title', 'Channel', 'Views', 'Likes',
              'Comments', 'Description', 'Licensed', 'official_video', 'Stream'],
              dtype='object')

In [6]: df.info
```

Out[6]:	<bound method DataFrame.info of	Unnamed: 0	Artist	\
0	0	Gorillaz		
1	1	Gorillaz		
2	2	Gorillaz		
3	3	Gorillaz		
4	4	Gorillaz		
...		
20713	20713	SICK LEGEND		
20714	20714	SICK LEGEND		
20715	20715	SICK LEGEND		
20716	20716	SICK LEGEND		
20717	20717	SICK LEGEND		
			Url_spotify	\
0		https://open.spotify.com/artist/3AA28KZvwAUcZu...		
1		https://open.spotify.com/artist/3AA28KZvwAUcZu...		
2		https://open.spotify.com/artist/3AA28KZvwAUcZu...		
3		https://open.spotify.com/artist/3AA28KZvwAUcZu...		
4		https://open.spotify.com/artist/3AA28KZvwAUcZu...		
...		...		
20713		https://open.spotify.com/artist/3EYY5FwDkHEYLw...		
20714		https://open.spotify.com/artist/3EYY5FwDkHEYLw...		
20715		https://open.spotify.com/artist/3EYY5FwDkHEYLw...		
20716		https://open.spotify.com/artist/3EYY5FwDkHEYLw...		
20717		https://open.spotify.com/artist/3EYY5FwDkHEYLw...		
			Track	\
0		Feel Good Inc.		
1		Rhinestone Eyes		
2	New Gold (feat. Tame Impala and Bootie Brown)			
3	On Melancholy Hill			
4	Clint Eastwood			
...		...		
20713	JUST DANCE HARDSTYLE			
20714	SET FIRE TO THE RAIN HARDSTYLE			
20715	OUTSIDE HARDSTYLE SPED UP			
20716	ONLY GIRL HARDSTYLE			
20717	MISS YOU HARDSTYLE			
			Album Album_type	\
0		Demon Days	album	
1		Plastic Beach	album	
2	New Gold (feat. Tame Impala and Bootie Brown)		single	
3	Plastic Beach		album	
4	Gorillaz		album	
...		
20713	JUST DANCE HARDSTYLE		single	
20714	SET FIRE TO THE RAIN HARDSTYLE		single	
20715	OUTSIDE HARDSTYLE SPED UP		single	
20716	ONLY GIRL HARDSTYLE		single	
20717	MISS YOU HARDSTYLE		single	
		Uri	Danceability	Energy
0	spotify:track:0d28khcov6AiegSCpG5TuT	0.818	0.705	6.0
1	spotify:track:1foMv2HQwfQ2vntFF9HFeG	0.676	0.703	8.0
2	spotify:track:64dLd6rVqDLtkXFYrEUHIU	0.695	0.923	1.0
3	spotify:track:0q6LuUqGLUiCPP1cbdwFs3	0.689	0.739	2.0
4	spotify:track:7yMiX7n9SBvadzoX8T5jzT	0.663	0.694	10.0
...
20713	spotify:track:0RtcKQGyI4hr8FgFH1TuYG	0.582	0.926	5.0
20714	spotify:track:3rHvPA8lUnPBkaLyPOc0VV	0.531	0.936	4.0
20715	spotify:track:4jk00YxPtPbhvHJE9N4ddv	0.443	0.830	4.0
20716	spotify:track:5EyErbpsugWliX006eTDex	0.417	0.767	9.0
20717	spotify:track:6l0n0jz1QpjcWeXo1oMm0k	0.498	0.938	6.0
		Uri_youtube		\
0		https://www.youtube.com/watch?v=HyHNuVaZJ-k		
1		https://www.youtube.com/watch?v=yYDmaexVHic		
2		https://www.youtube.com/watch?v=qJa-VFwPpYA		
3		https://www.youtube.com/watch?v=04mfKJWDSzI		
4		https://www.youtube.com/watch?v=1V_xRb0x9aw		
...		...		
20713		https://www.youtube.com/watch?v=5SHmKFKlNqI		
20714		https://www.youtube.com/watch?v=ocTH6KxllDQ		
20715		https://www.youtube.com/watch?v=5wFhE-HY0hg		
20716		https://www.youtube.com/watch?v=VMFLbFRNCn0		
20717		https://www.youtube.com/watch?v=zau0dckCFi0		
			Title	Channel
0		Gorillaz - Feel Good Inc. (Official Video)		Gorillaz
1		Gorillaz - Rhinestone Eyes [Storyboard Film] (...)		Gorillaz
2		Gorillaz - New Gold ft. Tame Impala & Bootie B...		Gorillaz
3		Gorillaz - On Melancholy Hill (Official Video)		Gorillaz
4		Gorillaz - Clint Eastwood (Official Video)		Gorillaz
...	
20713		JUST DANCE HARDSTYLE	SICK LEGEND - Topic	
20714		SET FIRE TO THE RAIN HARDSTYLE	SICK LEGEND - Topic	
20715		OUTSIDE HARDSTYLE SPED UP	SICK LEGEND - Topic	
20716		ONLY GIRL HARDSTYLE	SICK LEGEND - Topic	
20717		MISS YOU HARDSTYLE	SICK LEGEND - Topic	
		Views	Likes	Comments
0		693555221.0	6220896.0	169907.0
1		72011645.0	1079128.0	31003.0
2		8435055.0	282142.0	7399.0
3		211754952.0	1788577.0	55229.0
4		618480958.0	6197318.0	155930.0
...	
20713		71678.0	1113.0	0.0
20714		164741.0	2019.0	0.0
20715		35646.0	329.0	0.0
20716		6533.0	88.0	0.0
20717		158697.0	2484.0	0.0
			Description	Licensed
0		Official HD Video for Gorillaz' fantastic trac...		True
1		The official video for Gorillaz - Rhinestone E...		True
2		Gorillaz - New Gold ft. Tame Impala & Bootie B...		True
3		Follow Gorillaz online:\nhhttp://gorillaz.com \...		True
4		The official music video for Gorillaz - Clint ...		True
...	
20713		Provided to YouTube by Routenote\n\nJUST DANCE...		True
20714		Provided to YouTube by Routenote\n\nSET FIRE T...		True
20715		Provided to YouTube by Routenote\n\nOUTSIDE HA...		True
20716		Provided to YouTube by Routenote\n\nONLY GIRL ...		True
20717		Provided to YouTube by Routenote\n\nMISS YOU H...		True
		official_video	Stream	
0		True	1.040235e+09	
1		True	3.100837e+08	
2		True	6.306347e+07	
3		True	4.346636e+08	
4		True	6.172597e+08	
...		
20713		True	9.227144e+06	
20714		True	1.089818e+07	

```
20715      True  6.226110e+06
20716      True  6.873961e+06
20717      True  5.695584e+06
```

[20718 rows x 28 columns]>

drop columns

```
In [8]: df.drop(columns = ['Unnamed: 0','Url_spotify','Uri','Url_youtube'],inplace = True)
# here is removing unwanted columns.

In [9]: df.columns # check cloumns.

Out[9]: Index(['Artist', 'Track', 'Album', 'Album_type', 'Danceability', 'Energy',
              'Key', 'Loudness', 'Speechiness', 'Acousticness', 'Instrumentalness',
              'Liveness', 'Valence', 'Tempo', 'Duration_ms', 'Title', 'Channel',
              'Views', 'Likes', 'Comments', 'Description', 'Licensed',
              'official_video', 'Stream'],
              dtype='object')

In [10]: df.isna().sum()

Out[10]: Artist      0
Track      0
Album      0
Album_type  0
Danceability  2
Energy      2
Key         2
Loudness    2
Speechiness  2
Acousticness 2
Instrumentalness 2
Liveness    2
Valence     2
Tempo       2
Duration_ms  2
Title       470
Channel     470
Views       470
Likes       541
Comments    569
Description  876
Licensed    470
official_video 470
Stream      576
dtype: int64
```

Fill Missing Values

```
In [12]: # filling the missing value with 0 in Likes & comments column
df["Likes"] = df["Likes"].fillna(0)
df["Comments"] = df["Comments"].fillna(0)

In [13]: df.isnull().sum()

Out[13]: Artist      0
Track      0
Album      0
Album_type  0
Danceability  2
Energy      2
Key         2
Loudness    2
Speechiness  2
Acousticness 2
Instrumentalness 2
Liveness    2
Valence     2
Tempo       2
Duration_ms  2
Title       470
Channel     470
Views       470
Likes       0
Comments    0
Description  876
Licensed    470
official_video 470
Stream      576
dtype: int64
```

Drop Null Values

```
In [15]: df.dropna(inplace = True)

In [16]: df.isnull().sum() # to check the count of missing values in each column

Out[16]: Artist      0
Track      0
Album      0
Album_type  0
Danceability  0
Energy      0
Key         0
Loudness    0
Speechiness  0
Acousticness 0
Instrumentalness 0
Liveness    0
Valence     0
Tempo       0
Duration_ms  0
Title       0
Channel     0
Views       0
Likes       0
Comments    0
Description  0
Licensed    0
official_video 0
Stream      0
dtype: int64

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

```
<class 'pandas.core.frame.DataFrame'>
Index: 19298 entries, 0 to 20717
Data columns (total 24 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Artist                 19298 non-null object
1   Track                  19298 non-null object
2   Album                  19298 non-null object
3   Album_type             19298 non-null object
4   Danceability           19298 non-null float64
5   Energy                 19298 non-null float64
6   Key                    19298 non-null float64
7   Loudness               19298 non-null float64
8   Speechiness            19298 non-null float64
9   Acousticness           19298 non-null float64
10  Instrumentalness       19298 non-null float64
11  Liveness               19298 non-null float64
12  Valence                 19298 non-null float64
13  Tempo                  19298 non-null float64
14  Duration_ms            19298 non-null float64
15  Title                  19298 non-null object
16  Channel                19298 non-null object
17  Views                  19298 non-null float64
18  Likes                  19298 non-null float64
19  Comments               19298 non-null float64
20  Description            19298 non-null object
21  Licensed               19298 non-null object
22  official_video         19298 non-null object
23  Stream                 19298 non-null float64
dtypes: float64(15), object(9)
memory usage: 3.7+ MB
```

In []:

Q.1) Top 10 artists - With The Highest Views On Youtube ?

In [19]:

df.tail()

Out[19]:

	Artist	Track	Album	Album_type	Danceability	Energy	Key	Loudness	Speechiness	Acousticness	...	Duration_ms	Title	Channel	Views	Likes	Comments	Descriptio
20713	SICK LEGEND	JUST DANCE HARDSTYLE	JUST DANCE HARDSTYLE	single	0.582	0.926	5.0	-6.344	0.0328	0.44800	...	94667.0	JUST DANCE HARDSTYLE	SICK LEGEND - Topic	71678.0	1113.0	0.0	Provided to YouTube by Routenote\n\nJUS DANCE
20714	SICK LEGEND	SET FIRE TO THE RAIN HARDSTYLE	SET FIRE TO THE RAIN HARDSTYLE	single	0.531	0.936	4.0	-1.786	0.1370	0.02800	...	150857.0	SET FIRE TO THE RAIN HARDSTYLE	SICK LEGEND - Topic	164741.0	2019.0	0.0	Provided to YouTube by Routenote\n\nSt FIRE T
20715	SICK LEGEND	OUTSIDE HARDSTYLE SPED UP	OUTSIDE HARDSTYLE SPED UP	single	0.443	0.830	4.0	-4.679	0.0647	0.02430	...	136842.0	OUTSIDE HARDSTYLE SPED UP	SICK LEGEND - Topic	35646.0	329.0	0.0	Provided to YouTube t Routenote\n\nOUTSIC HA
20716	SICK LEGEND	ONLY GIRL HARDSTYLE	ONLY GIRL HARDSTYLE	single	0.417	0.767	9.0	-4.004	0.4190	0.35600	...	108387.0	ONLY GIRL HARDSTYLE	SICK LEGEND - Topic	6533.0	88.0	0.0	Provided to YouTube by Routenote\n\nnONI GIRL
20717	SICK LEGEND	MISS YOU HARDSTYLE	MISS YOU HARDSTYLE	single	0.498	0.938	6.0	-4.543	0.1070	0.00277	...	181500.0	MISS YOU HARDSTYLE	SICK LEGEND - Topic	158697.0	2484.0	0.0	Provided to YouTube by Routenote\n\nnMIS YOU H

5 rows × 24 columns

In [20]:

```
top_artists = (
    df.groupby("Artist")["Views"]
      .sum()
      .sort_values(ascending = False)
      .head(10)
)
print(top_artists)  # here is first question results
```

Artist	
Ed Sheeran	1.546021e+10
CoComelon	1.460167e+10
Katy Perry	1.312063e+10
Charlie Puth	1.216759e+10
Luis Fonsi	1.162811e+10
Justin Bieber	1.099079e+10
Daddy Yankee	1.086828e+10
Bruno Mars	1.023184e+10
Macklemore & Ryan Lewis	1.012206e+10
Coldplay	9.997278e+09

Name: Views, dtype: float64

In []:

Top 10 Tracks - with The Highest Streams On Spotify

In [22]:

df.head(1)

Out[22]:

	Artist	Track	Album	Album_type	Danceability	Energy	Key	Loudness	Speechiness	Acousticness	...	Duration_ms	Title	Channel	Views	Likes	Comments	Description	Licensed	official_v
0	Gorillaz	Feel Good Inc.	Demon Days	album	0.818	0.705	6.0	-6.679	0.177	0.00836	...	222640.0	Gorillaz - Feel Good Inc. (Official Video)	Gorillaz	693555221.0	6220896.0	169907.0	Official HD Video for Gorillaz' fantastic trac...	True	

1 rows × 24 columns

In [23]:

```
most_stream_track = (
    df.groupby("Track", as_index=False)["Stream"]
      .sum()
      .sort_values(by="Stream", ascending=False)
      .head(10)
)

print(most_stream_track)  # Print the result
```

	Track	Stream
2703	Closer	5.465365e+09
2365	Can't Hold Us (feat. Ray Dalton)	5.225629e+09
13164	Sunflower - Spider-Man: Into the Spider-Verse	5.076660e+09
5575	Happier	4.757615e+09
11842	STAY (with Justin Bieber)	4.731555e+09
12239	Señorita	4.672440e+09
13963	The Middle	4.566883e+09
3964	Eastside (with Halsey & Khalid)	4.274888e+09
16384	lovely (with Khalid)	4.221148e+09
4242	Enemy (with JID) - from the series Arcane Leag...	4.182328e+09

In []:

Q.2A) 5 Tracks - with the Lowest Streams on Spotify?

```
In [25]: lowest_stream_track = (  
    df.groupby("Track", as_index = False)["Stream"]  
    .sum()  
    .sort_values(by = "Stream", ascending = True)  
    .head(5)  
)  
print(lowest_stream_track)
```

	Track	Stream
11321	Rainbows	6574.0
15389	We Play the Funk - Slynk Remix	7771.0
13623	Teil 9 - Sherlock Holmes und ein Brief von der...	8053.0
13608	Teil 10 - Sherlock Holmes und ein Brief von de...	8074.0
13614	Teil 5 - Sherlock Holmes und der blinde Bettle...	10306.0

In []:

Q.3) What are the most common Album Types on Spotify? How many tracks belong to each album type?

```
In [27]: df.head(1)
```

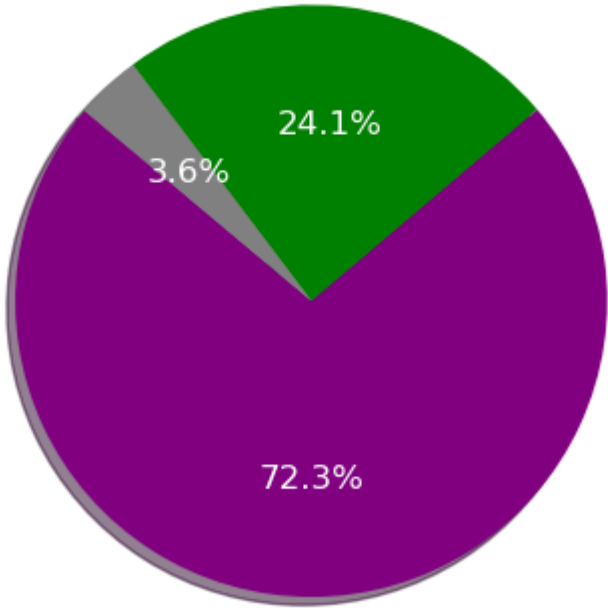
	Artist	Track	Album	Album_type	Danceability	Energy	Key	Loudness	Speechiness	Acousticness	...	Duration_ms	Title	Channel	Views	Likes	Comments	Description	Licensed	official_v
0	Gorillaz	Feel Good Inc.	Demon Days	album	0.818	0.705	6.0	-6.679	0.177	0.00836	...	222640.0	Gorillaz - Feel Good Inc. (Official Video)	Gorillaz	693555221.0	6220896.0	169907.0	Official HD Video for Gorillaz' fantastic trac...	True	
1 rows × 24 columns																				

```
In [28]: a_type = df["Album_type"].value_counts()  
a_type
```

```
Out[28]: Album_type  
album      13952  
single      4653  
compilation    693  
Name: count, dtype: int64
```

```
In [29]: labels = ["album","single","compilation"]  
sizes = [13955, 4653, 693]  
plt.pie(  
    sizes,  
    labels = labels,  
    autopct = "%1.1f%",  
    startangle = 140,  
    shadow = True,  
    colors = ["purple", "green", "grey"],  
    textprops = {"color": "white", "fontsize": 12}  
)  
  
plt.title("Most Common Album On Spotify",color = "blue")  
plt.show()
```

Most Common Album On Spotify



In []:

Q.4) How do the Average Views, Likes, and Comments are compared between different Album Types?

```
In [31]: avg_stats = df.groupby("Album_type")[["Views","Likes","Comments"]].mean()  
print(avg_stats)
```

Album_type	Views	Likes	Comments
album	1.014359e+08	672593.609948	29032.951548
compilation	8.500535e+07	544275.818182	18074.903319
single	8.480843e+07	722266.303890	27313.048571

```
In [32]: type(df)
```

```
Out[32]: pandas.core.frame.DataFrame
```

```
In [33]: # rest_index - To convert the index of a Series into a column to form a DataFrame
```

df = df.reset_index()
df

	index	Artist	Track	Album	Album_type	Danceability	Energy	Key	Loudness	Speechiness	...	Duration_ms	Title	Channel	Views	Likes	Comments	Descri
0	0	Gorillaz	Feel Good Inc.	Demon Days	album	0.818	0.705	6.0	-6.679	0.1770	...	222640.0	Gorillaz - Feel Good Inc. (Official Video)	Gorillaz	693555221.0	6220896.0	169907.0	Official HD Video of Gorillaz' fantastic
1	1	Gorillaz	Rhinestone Eyes	Plastic Beach	album	0.676	0.703	8.0	-5.815	0.0302	...	200173.0	Gorillaz - Rhinestone Eyes [Storyboard Film] (...)	Gorillaz	72011645.0	1079128.0	31003.0	The official video of Gorillaz - Rhinestone Eyes
2	2	Gorillaz	New Gold (feat. Tame Impala and Bootie Brown)	New Gold (feat. Tame Impala and Bootie Brown)	single	0.695	0.923	1.0	-3.930	0.0522	...	215150.0	Gorillaz - New Gold ft. Tame Impala & Bootie Brown	Gorillaz	8435055.0	282142.0	7399.0	Gorillaz - New Gold ft. Tame Impala & Bootie Brown
3	3	Gorillaz	On Melancholy Hill	Plastic Beach	album	0.689	0.739	2.0	-5.810	0.0260	...	233867.0	Gorillaz - On Melancholy Hill (Official Video)	Gorillaz	211754952.0	1788577.0	55229.0	Follow Gorillaz online:\nhttp://gorillaz.com
4	4	Gorillaz	Clint Eastwood	Gorillaz	album	0.663	0.694	10.0	-8.627	0.1710	...	340920.0	Gorillaz - Clint Eastwood (Official Video)	Gorillaz	618480958.0	6197318.0	155930.0	The official music video of Gorillaz - Clint Eastwood
...
19293	20713	SICK LEGEND	JUST DANCE HARDSTYLE	JUST DANCE HARDSTYLE	single	0.582	0.926	5.0	-6.344	0.0328	...	94667.0	JUST DANCE HARDSTYLE	SICK LEGEND - Topic	71678.0	1113.0	0.0	Provided to YouTube by Sick Legend - Topic Routenote\n\nDAI
19294	20714	SICK LEGEND	SET FIRE TO THE RAIN HARDSTYLE	SET FIRE TO THE RAIN HARDSTYLE	single	0.531	0.936	4.0	-1.786	0.1370	...	150857.0	SET FIRE TO THE RAIN HARDSTYLE	SICK LEGEND - Topic	164741.0	2019.0	0.0	Provided to YouTube by Sick Legend - Topic Routenote\n\nnSET FIRE TO THE RAIN HARDSTYLE
19295	20715	SICK LEGEND	OUTSIDE HARDSTYLE SPED UP	OUTSIDE HARDSTYLE SPED UP	single	0.443	0.830	4.0	-4.679	0.0647	...	136842.0	OUTSIDE HARDSTYLE SPED UP	SICK LEGEND - Topic	35646.0	329.0	0.0	Provided to YouTube by Sick Legend - Topic Routenote\n\nnOUTSIDE HARDSTYLE SPED UP
19296	20716	SICK LEGEND	ONLY GIRL HARDSTYLE	ONLY GIRL HARDSTYLE	single	0.417	0.767	9.0	-4.004	0.4190	...	108387.0	ONLY GIRL HARDSTYLE	SICK LEGEND - Topic	6533.0	88.0	0.0	Provided to YouTube by Sick Legend - Topic Routenote\n\nnONLY GIRL HARDSTYLE
19297	20717	SICK LEGEND	MISS YOU HARDSTYLE	MISS YOU HARDSTYLE	single	0.498	0.938	6.0	-4.543	0.1070	...	181500.0	MISS YOU HARDSTYLE	SICK LEGEND - Topic	158697.0	2484.0	0.0	Provided to YouTube by Sick Legend - Topic Routenote\n\nnMISS YOU HARDSTYLE

19298 rows × 25 columns

In [34]:

Melt - unpivot a Dataframe
df_melted = pd.melt(
 df,
 id_vars = ["Album_type"], # grouping key
 value_vars = ["Views", "Likes", "Comments"], # only 3 columns
 var_name = "Attribute",
 value_name = "Total"
)
df_melted

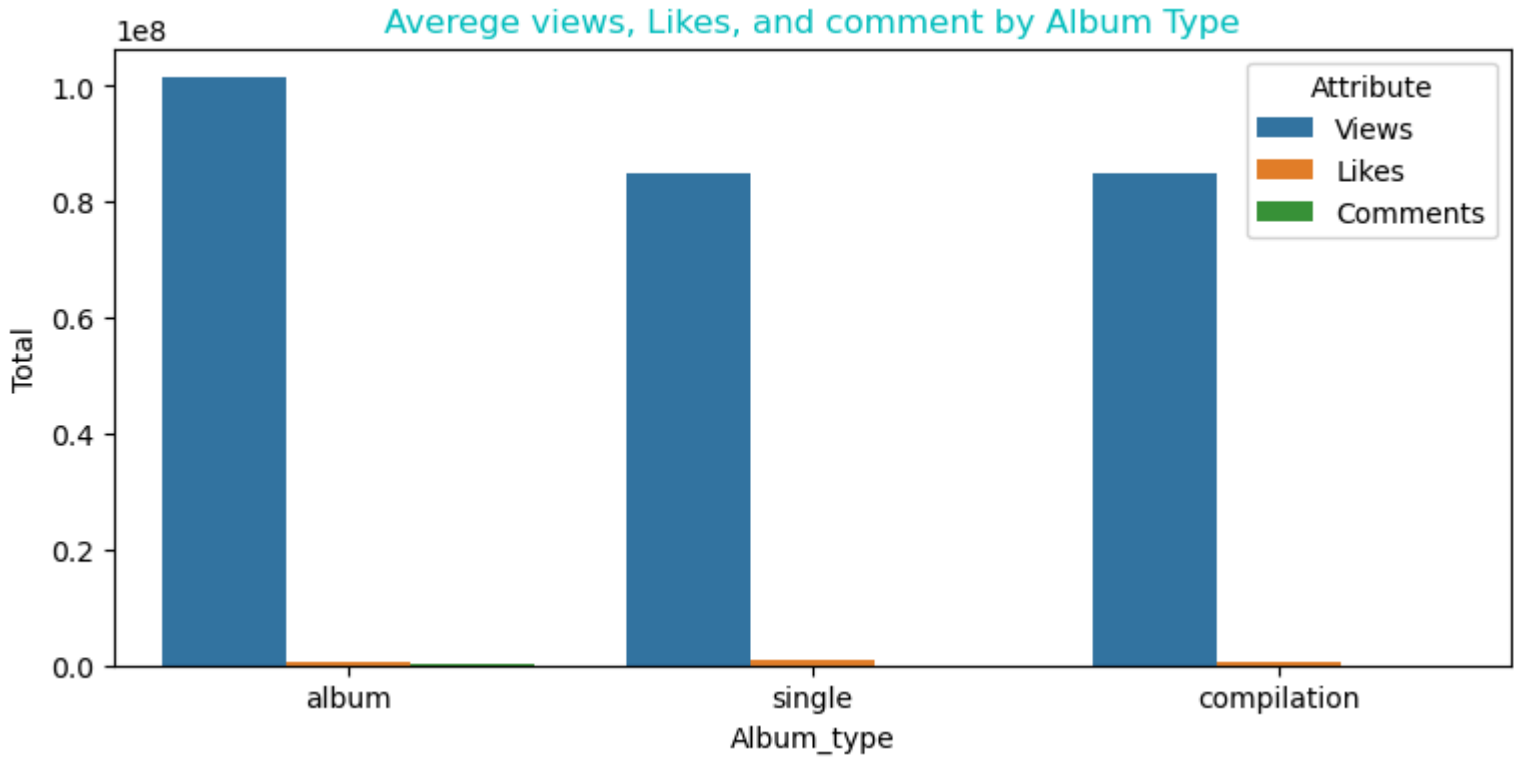
	Album_type	Attribute	Total
0	album	Views	693555221.0
1	album	Views	72011645.0
2	single	Views	8435055.0
3	album	Views	211754952.0
4	album	Views	618480958.0
...
57889	single	Comments	0.0
57890	single	Comments	0.0
57891	single	Comments	0.0
57892	single	Comments	0.0
57893	single	Comments	0.0

57894 rows × 3 columns

In [35]:

draw the bar plot
plt.figure(figsize = (9,4))
sns.barplot(

 x = "Album_type",
 y = "Total",
 hue = "Attribute",
 data = df_melted,
 errorbar = None
)
plt.title("Averege views, Likes, and comment by Album Type",color = "c")
plt.show()



In []:

Q.5) Top 5 YouTube Channels - based on the Views?

In [37]: `df.head(1)`

Out[37]:

	index	Artist	Track	Album	Album_type	Danceability	Energy	Key	Loudness	Speechiness	...	Duration_ms	Title	Channel	Views	Likes	Comments	Description	Licensed	official_video	
0	0	Gorillaz	Feel Good Inc.	Demon Days	album	0.818	0.705	6.0	-6.679	0.177	...	222640.0	Gorillaz - Feel Good Inc. (Official Video)	Gorillaz	693555221.0	6220896.0	169907.0	Official HD Video for Gorillaz' fantastic trac...	True	True	1

1 rows × 25 columns



In [38]: `channel_views = df.groupby("Channel")["Views"].sum().sort_values(ascending = False).head()`

channel_views

Out[38]:

Channel	Views
T-Series	4.796159e+10
LuisFonsiVEVO	2.170252e+10
CalvinHarrisVEVO	1.828114e+10
Macklemore	1.700341e+10
Ed Sheeran	1.691822e+10

Name: Views, dtype: float64

In [39]: `channel_views = channel_views.reset_index()`

channel_views.head(10)

Out[39]:

	Channel	Views
0	T-Series	4.796159e+10
1	LuisFonsiVEVO	2.170252e+10
2	CalvinHarrisVEVO	1.828114e+10
3	Macklemore	1.700341e+10
4	Ed Sheeran	1.691822e+10

In [40]: `type(channel_views)`

Out[40]: `pandas.core.frame.DataFrame`

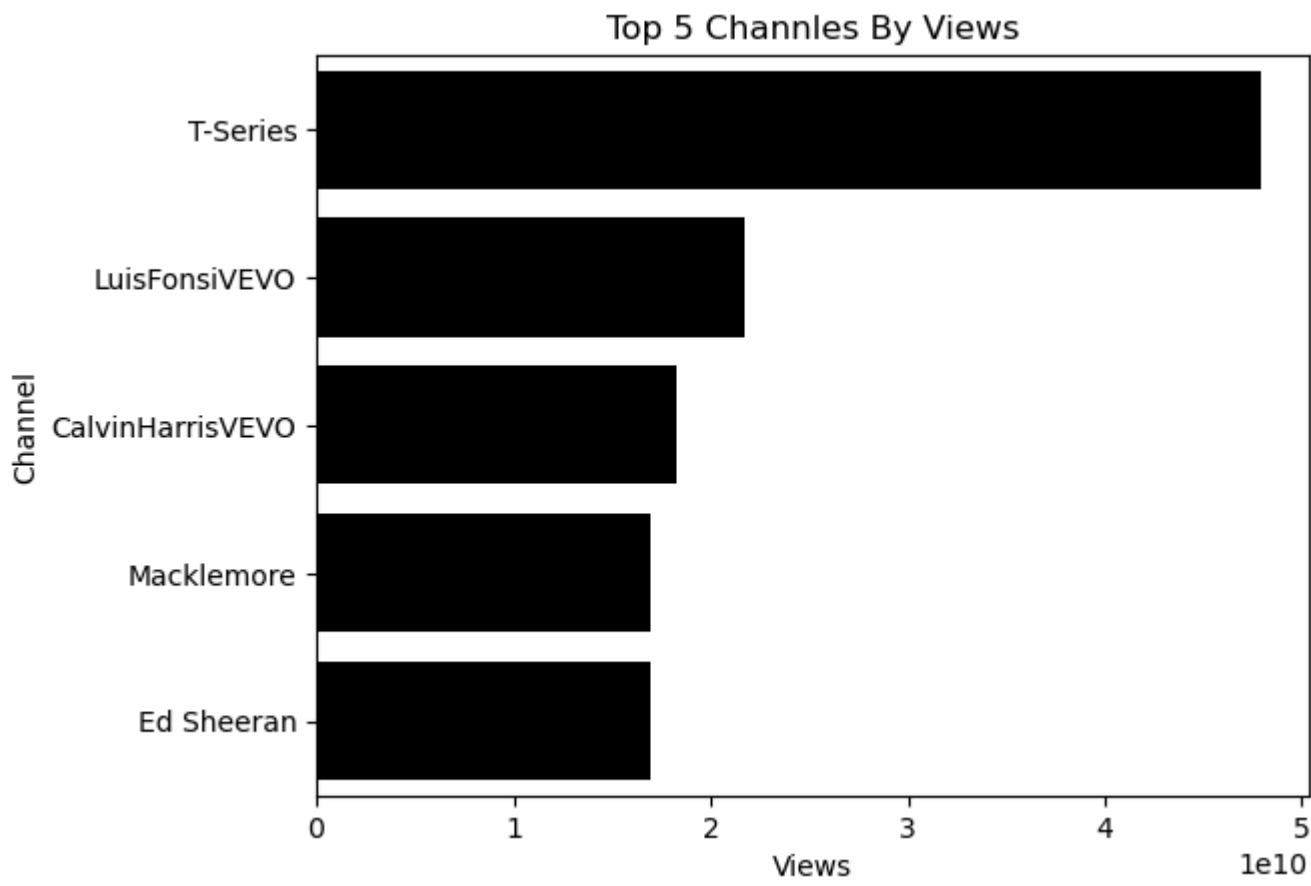
In [41]: `sns.barplot(x = "Views", y = "Channel", data = channel_views, color = "black")`

`plt.title("Top 5 Channles By Views")`

`plt.xlabel("Views")`

`plt.ylabel("Channel")`

`plt.show()`



In []:

Q.6) The Top Most Track - Based on views ?

In [43]: `df.head (3)`

Out[43]:

	index	Artist	Track	Album	Album_type	Danceability	Energy	Key	Loudness	Speechiness	...	Duration_ms	Title	Channel	Views	Likes	Comments	Description	Licensed	official_
0	0	Gorillaz	Feel Good Inc.	Demon Days	album	0.818	0.705	6.0	-6.679	0.1770	...	222640.0	Gorillaz - Feel Good Inc. (Official Video)	Gorillaz	693555221.0	6220896.0	169907.0	Official HD Video for Gorillaz' fantastic trac...	True	
1	1	Gorillaz	Rhinestone Eyes	Plastic Beach	album	0.676	0.703	8.0	-5.815	0.0302	...	200173.0	Gorillaz - Rhinestone Eyes [Storyboard Film] (...)	Gorillaz	72011645.0	1079128.0	31003.0	The official video for Gorillaz - Rhinestone E...	True	
2	2	Gorillaz	New Gold (feat. Tame Impala and Bootie Brown)	New Gold (feat. Tame Impala and Bootie Brown)	single	0.695	0.923	1.0	-3.930	0.0522	...	215150.0	Gorillaz - New Gold ft. Tame Impala & Bootie B...	Gorillaz	8435055.0	282142.0	7399.0	Gorillaz - New Gold ft. Tame Impala & Bootie B...	True	

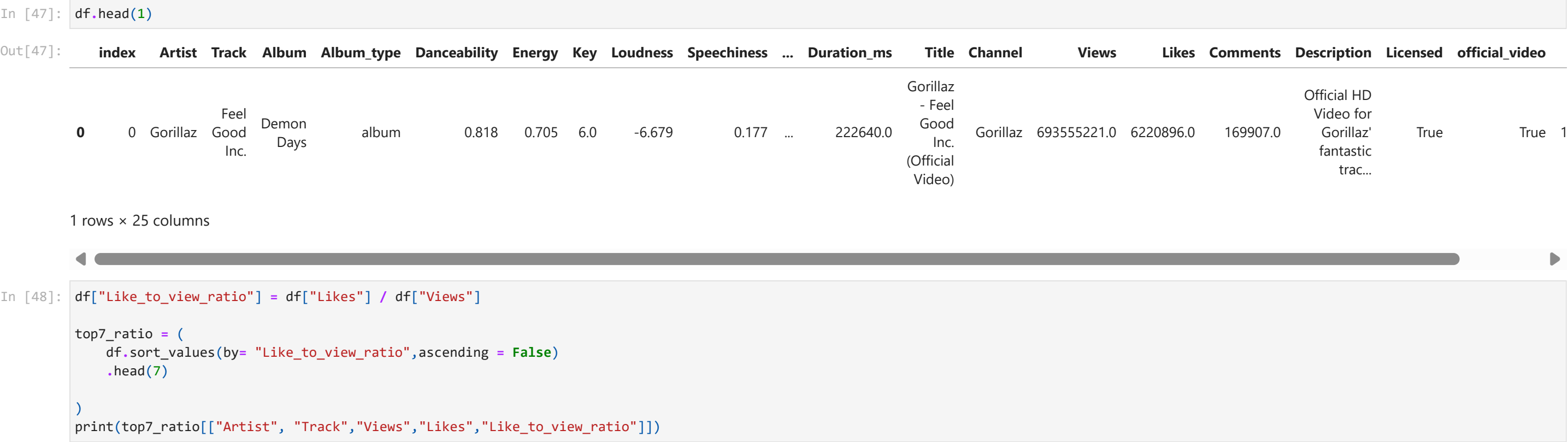
3 rows × 25 columns



Top 5 Tracks by Views



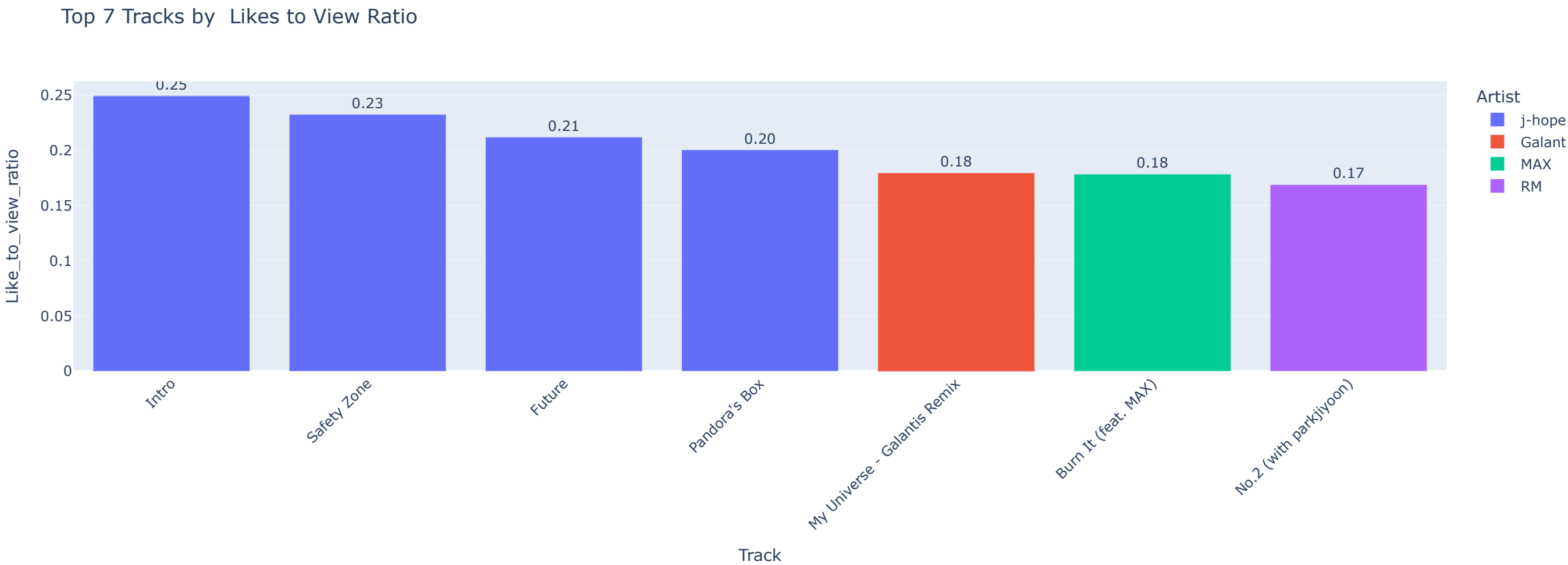
Q.7) Which Top 7 Tracks Have The Highest Like_to-View Ratio on Youtube



	Artist	Track	Views	Likes	\
18622	j-hope	Intro	954081.0	237761.0	
18623	j-hope	Safety Zone	1952637.0	453910.0	
18621	j-hope	Future	1180522.0	250116.0	
18625	j-hope	Pandora's Box	1265231.0	253702.0	
15108	Galantis	My Universe - Galantis Remix	2067753.0	371437.0	
13933	MAX	Burn It (feat. MAX)	1054438.0	188244.0	
7460	RM	No.2 (with parkjiyon)	2050047.0	346440.0	

	Like_to_view_ratio
18622	0.249204
18623	0.232460
18621	0.211869
18625	0.200518
15108	0.179633
13933	0.178525
7460	0.168991

```
In [49]: fig = px.bar(
        top7_ratio,
        x="Track",
        y="Like_to_view_ratio",
        color="Artist",
        text="Like_to_view_ratio",
        title="Top 7 Tracks by Likes to View Ratio"
    )
fig.update_traces(texttemplate="%{text:.2f}",textposition="outside")
fig.update_layout(xaxis_tickangle = -45)
fig.show()
```



```
In [ ]:
```

Q.7.A) Which Top 3 Tracks have the lowest Like-to-View ratio on YouTube?

```
In [51]: df.head()
```

```
Out[51]:
```

	index	Artist	Track	Album	Album_type	Danceability	Energy	Key	Loudness	Speechiness	...	Title	Channel	Views	Likes	Comments	Description	Licensed	official
0	0	Gorillaz	Feel Good Inc.	Demon Days	album	0.818	0.705	6.0	-6.679	0.1770	...	Gorillaz - Feel Good Inc. (Official Video)	Gorillaz	693555221.0	6220896.0	169907.0	Official HD Video for Gorillaz' fantastic trac...	True	
1	1	Gorillaz	Rhinestone Eyes	Plastic Beach	album	0.676	0.703	8.0	-5.815	0.0302	...	Gorillaz - Rhinestone Eyes [Storyboard Film] (...)	Gorillaz	72011645.0	1079128.0	31003.0	The official video for Gorillaz - Rhinestone E...	True	
2	2	Gorillaz	New Gold (feat. Tame Impala and Bootie Brown)	New Gold (feat. Tame Impala and Bootie Brown)	single	0.695	0.923	1.0	-3.930	0.0522	...	Gorillaz - New Gold ft. Tame Impala & Bootie B...	Gorillaz	8435055.0	282142.0	7399.0	Gorillaz - New Gold ft. Tame Impala & Bootie B...	True	
3	3	Gorillaz	On Melancholy Hill	Plastic Beach	album	0.689	0.739	2.0	-5.810	0.0260	...	Gorillaz - On Melancholy Hill (Official Video)	Gorillaz	211754952.0	1788577.0	55229.0	Follow Gorillaz online:\nhttp://gorillaz.com \...	True	
4	4	Gorillaz	Clint Eastwood	Gorillaz	album	0.663	0.694	10.0	-8.627	0.1710	...	Gorillaz - Clint Eastwood (Official Video)	Gorillaz	618480958.0	6197318.0	155930.0	The official music video for Gorillaz - Clint ...	True	

5 rows × 26 columns

```
In [52]: # Calculate Like-to-View ratio
df["Like_to_View"] = df["Likes"] / df["Views"]

lowest_ratio_tracks = df.sort_values("Like_to_View").head(3)

print(lowest_ratio_tracks[["Track", "Views", "Likes", "Like_to_View"]])
```

	Track	Views	Likes	Like_to_View
9223	Persona Ideal - Me Tengo Que Ir	13627867.0	0.0	0.0
5147	April In Paris	216572.0	0.0	0.0
5062	Calaverada	39830.0	0.0	0.0

Q.8) Top Albums having the Tracks with Maximum Danceability ?

```
In [54]: df.head(1)
```

Out[54]:

	index	Artist	Track	Album	Album_type	Danceability	Energy	Key	Loudness	Speechiness	...	Channel	Views	Likes	Comments	Description	Licensed	official_video	Stream	Like_to_vi
0	0	Gorillaz	Feel Good Inc.	Demon Days	album	0.818	0.705	6.0	-6.679	0.177	...	Gorillaz	693555221.0	6220896.0	169907.0	Official HD Video for Gorillaz' fantastic trac...	True	True	1.040235e+09	

1 rows × 27 columns

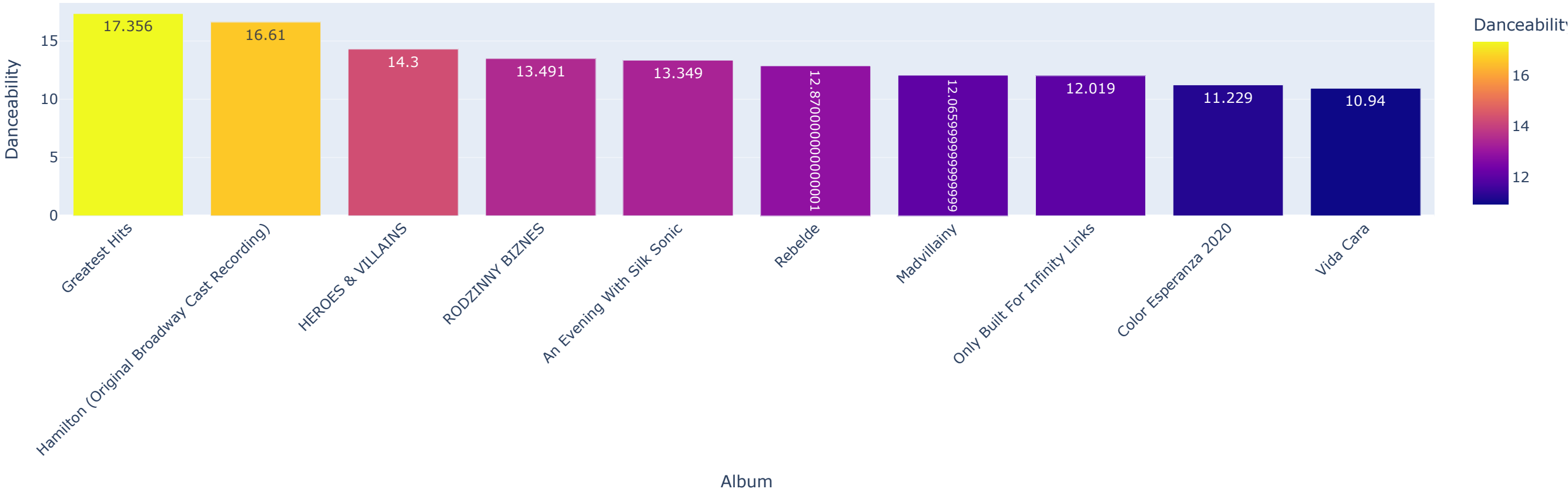
In [55]:

```
tr_danceability = (
    df.groupby("Album")["Danceability"]
      .sum()
      .sort_values(ascending = False)
      .reset_index()
      .head(10)
)
print(tr_danceability)

fig = px.bar(
    tr_danceability,
    x="Album",
    y="Danceability",
    title="Top 10 albums by Danceability Sum",
    color="Danceability",
    text="Danceability"
)
fig.update_layout(xaxis_tickangle= -45)
fig.show()
```

	Album	Danceability
0	Greatest Hits	17.356
1	Hamilton (Original Broadway Cast Recording)	16.610
2	HEROES & VILLAINS	14.300
3	RODZINNY BIZNES	13.491
4	An Evening With Silk Sonic	13.349
5	Rebelde	12.870
6	Madvillainy	12.066
7	Only Built For Infinity Links	12.019
8	Color Esperanza 2020	11.229
9	Vida Cara	10.940

Top 10 albums by Danceability Sum



Q.9) What is the Correlation between Views, Likes, Comments, and Stream?

In [57]:

df.head(1)

Out[57]:

	index	Artist	Track	Album	Album_type	Danceability	Energy	Key	Loudness	Speechiness	...	Channel	Views	Likes	Comments	Description	Licensed	official_video	Stream	Like_to_vi
0	0	Gorillaz	Feel Good Inc.	Demon Days	album	0.818	0.705	6.0	-6.679	0.177	...	Gorillaz	693555221.0	6220896.0	169907.0	Official HD Video for Gorillaz' fantastic trac...	True	True	1.040235e+09	

1 rows × 27 columns

In [58]:

```
df_vlcs = df[["Views", "Likes", "Comments", "Stream"]]
print(df_vlcs)
```

	Views	Likes	Comments	Stream
0	693555221.0	6220896.0	169907.0	1.040235e+09
1	72011645.0	1079128.0	31003.0	3.100837e+08
2	8435055.0	282142.0	7399.0	6.306347e+07
3	211754952.0	1788577.0	55229.0	4.346636e+08
4	618480958.0	6197318.0	155930.0	6.172597e+08
...
19293	71678.0	1113.0	0.0	9.227144e+06
19294	164741.0	2019.0	0.0	1.089818e+07
19295	35646.0	329.0	0.0	6.226110e+06
19296	6533.0	88.0	0.0	6.873961e+06
19297	158697.0	2484.0	0.0	5.695584e+06

[19298 rows x 4 columns]

In [59]:

df_vlcs.corr()

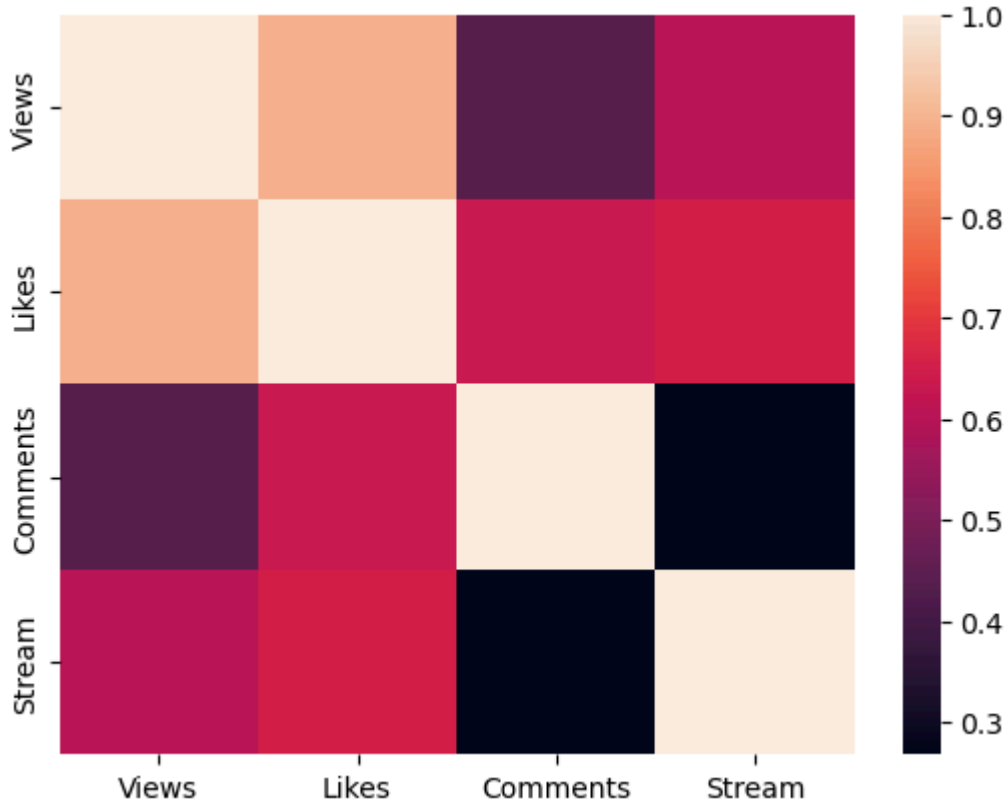
Out[59]:

	Views	Likes	Comments	Stream
Views	1.000000	0.891695	0.431077	0.603510
Likes	0.891695	1.000000	0.631035	0.655808
Comments	0.431077	0.631035	1.000000	0.267833
Stream	0.603510	0.655808	0.267833	1.000000

In [60]:

sns.heatmap(df_vlcs.corr())

Out[60]: <Axes: >



In []: