

Spotify Data analysis

August 2, 2023

SPOTIFY DATA ANALYSIS

BY DEEPAK YADAV(dy479958@gmail.com)

```
[ ]: # IMPORTING REQUIRED LIBRARIES
```

```
[1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
%matplotlib inline
```

```
[2]: #LOADING OUR DATASET
spotify_data= pd.read_csv("D:\Spotify_Youtube.csv")
#basic view of the dataset
spotify_data.head()
```

```
[2]: Unnamed: 0    Artist                                Url_spotify \
0          0  Gorillaz  https://open.spotify.com/artist/3AA28KZvwAUcZu...
1          1  Gorillaz  https://open.spotify.com/artist/3AA28KZvwAUcZu...
2          2  Gorillaz  https://open.spotify.com/artist/3AA28KZvwAUcZu...
3          3  Gorillaz  https://open.spotify.com/artist/3AA28KZvwAUcZu...
4          4  Gorillaz  https://open.spotify.com/artist/3AA28KZvwAUcZu...
```

```
                                Track \
0                                Feel Good Inc.
1                                Rhinestone Eyes
2  New Gold (feat. Tame Impala and Bootie Brown)
3                                On Melancholy Hill
4                                Clint Eastwood
```

```
                                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
```

| | Uri | Danceability | Energy | Key | ... | \ |
|---|--------------------------------------|--------------|--------|------|-----|---|
| 0 | spotify:track:0d28kxcov6AiegSCpG5TuT | 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 | ... | |

| | Url_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_xRbOx9aw | |

| | Title | Channel | Views | \ |
|---|--|----------|-------------|---|
| 0 | Gorillaz - Feel Good Inc. (Official Video) | Gorillaz | 693555221.0 | |
| 1 | Gorillaz - Rhinestone Eyes [Storyboard Film] (...) | Gorillaz | 72011645.0 | |
| 2 | Gorillaz - New Gold ft. Tame Impala & Bootie B... | Gorillaz | 8435055.0 | |
| 3 | Gorillaz - On Melancholy Hill (Official Video) | Gorillaz | 211754952.0 | |
| 4 | Gorillaz - Clint Eastwood (Official Video) | Gorillaz | 618480958.0 | |

| | Likes | Comments | Description | \ |
|---|-----------|----------|---|---|
| 0 | 6220896.0 | 169907.0 | Official HD Video for Gorillaz' fantastic trac... | |
| 1 | 1079128.0 | 31003.0 | The official video for Gorillaz - Rhinestone E... | |
| 2 | 282142.0 | 7399.0 | Gorillaz - New Gold ft. Tame Impala & Bootie B... | |
| 3 | 1788577.0 | 55229.0 | Follow Gorillaz online:\nhttp://gorillaz.com \\\n | |
| 4 | 6197318.0 | 155930.0 | The official music video for Gorillaz - Clint ... | |

| | Licensed | official_video | Stream |
|---|----------|----------------|--------------|
| 0 | True | True | 1.040235e+09 |
| 1 | True | True | 3.100837e+08 |
| 2 | True | True | 6.306347e+07 |
| 3 | True | True | 4.346636e+08 |
| 4 | True | True | 6.172597e+08 |

[5 rows x 28 columns]

```
[3]: #Information about the dataset
spotify_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20718 entries, 0 to 20717
Data columns (total 28 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Unnamed: 0            20718 non-null  int64
1   Artist                20718 non-null  object
```

```

2  Url_spotify      20718 non-null object
3  Track           20718 non-null object
4  Album           20718 non-null object
5  Album_type      20718 non-null object
6  Uri             20718 non-null object
7  Danceability    20716 non-null float64
8  Energy          20716 non-null float64
9  Key             20716 non-null float64
10 Loudness        20716 non-null float64
11 Speechiness     20716 non-null float64
12 Acousticness    20716 non-null float64
13 Instrumentalness 20716 non-null float64
14 Liveness        20716 non-null float64
15 Valence         20716 non-null float64
16 Tempo          20716 non-null float64
17 Duration_ms     20716 non-null float64
18 Url_youtube     20248 non-null object
19 Title           20248 non-null object
20 Channel         20248 non-null object
21 Views           20248 non-null float64
22 Likes           20177 non-null float64
23 Comments        20149 non-null float64
24 Description     19842 non-null object
25 Licensed        20248 non-null object
26 official_video  20248 non-null object
27 Stream          20142 non-null float64
dtypes: float64(15), int64(1), object(12)
memory usage: 4.4+ MB

```

```
[4]: spotify_data.columns
```

```
[4]: 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')
```

```
[6]: #Cleaning unnecessary data
#removing unusable columns from dataset
df=spotify_data.drop(['Unnamed: 0', 'Url_spotify', 'Uri', 'Url_youtube', 'Title'],axis=1)
df.head()
```

```
[6]:      Artist                                     Track \
0  Gorillaz                                     Feel Good Inc.
1  Gorillaz                                     Rhinestone Eyes
2  Gorillaz  New Gold (feat. Tame Impala and Bootie Brown)
```

| | | |
|---|----------|--------------------|
| 3 | Gorillaz | On Melancholy Hill |
| 4 | Gorillaz | Clint Eastwood |

| | Album | Album_type | Danceability | \ |
|---|---|------------|--------------|---|
| 0 | Demon Days | album | 0.818 | |
| 1 | Plastic Beach | album | 0.676 | |
| 2 | New Gold (feat. Tame Impala and Bootie Brown) | single | 0.695 | |
| 3 | Plastic Beach | album | 0.689 | |
| 4 | Gorillaz | album | 0.663 | |

| | Energy | Key | Loudness | Speechiness | Acousticness | ... | Tempo | \ |
|---|--------|------|----------|-------------|--------------|-----|---------|---|
| 0 | 0.705 | 6.0 | -6.679 | 0.1770 | 0.008360 | ... | 138.559 | |
| 1 | 0.703 | 8.0 | -5.815 | 0.0302 | 0.086900 | ... | 92.761 | |
| 2 | 0.923 | 1.0 | -3.930 | 0.0522 | 0.042500 | ... | 108.014 | |
| 3 | 0.739 | 2.0 | -5.810 | 0.0260 | 0.000015 | ... | 120.423 | |
| 4 | 0.694 | 10.0 | -8.627 | 0.1710 | 0.025300 | ... | 167.953 | |

| | Duration_ms | Channel | Views | Likes | Comments | \ |
|---|-------------|----------|-------------|-----------|----------|---|
| 0 | 222640.0 | Gorillaz | 693555221.0 | 6220896.0 | 169907.0 | |
| 1 | 200173.0 | Gorillaz | 72011645.0 | 1079128.0 | 31003.0 | |
| 2 | 215150.0 | Gorillaz | 8435055.0 | 282142.0 | 7399.0 | |
| 3 | 233867.0 | Gorillaz | 211754952.0 | 1788577.0 | 55229.0 | |
| 4 | 340920.0 | Gorillaz | 618480958.0 | 6197318.0 | 155930.0 | |

| | 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 | |

| | 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 |

[5 rows x 23 columns]

```
[11]: #checking null values
df.isnull().sum()
```

```
[11]: 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 |
| Channel | 470 |
| Views | 470 |
| Likes | 541 |
| Comments | 569 |
| Description | 876 |
| Licensed | 470 |
| official_video | 470 |
| Stream | 576 |

dtype: int64

```
[12]: #Removing all null values from the columns
df.dropna(inplace=True)
```

```
[15]: #check again for null data
df.isnull().sum()
```

```
[15]: 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
Channel     0
Views       0
Likes       0
Comments    0
Description  0
```

```

Licensed          0
official_video    0
Stream            0
dtype: int64

```

```

[20]: #changing data types of some columns
df['Views']=df['Views'].astype('int')
df['Likes']=df['Likes'].astype('int')
df['Comments']=df['Comments'].astype('int')

```

```

[21]: #Visualization
#collecting top 10 artist whose music got most view
view=df.groupby('Artist')['Views'].agg('sum').reset_index().
    ↪sort_values(by='Views',ascending=False).head(10)
view

```

```

[21]:

```

| | Artist | Views |
|------|-------------------------|--------------|
| 263 | Bruno Mars | 1.023184e+10 |
| 1163 | Macklemore & Ryan Lewis | 1.012206e+10 |
| 375 | Coldplay | 9.997278e+09 |
| 149 | BLACKPINK | 9.392962e+09 |
| 780 | J Balvin | 8.464878e+09 |
| 1347 | Nicki Minaj | 8.017064e+09 |
| 512 | Earth, Wind & Fire | 6.414077e+09 |
| 1470 | Post Malone | 6.394420e+09 |
| 1390 | One Direction | 6.337602e+09 |
| 1349 | Nicky Jam | 6.321868e+09 |

```

[22]: #Barplot using plotly
view_visual=px.bar(view,x='Artist',y='Views',color='Artist',title='Top 10_
    ↪Viewed Artist')
view_visual.show()

```

```

[23]: # Collecting top 10 liked Artist
Like = df.groupby(['Artist'])['Likes'].sum().reset_index().
    ↪sort_values('Likes',ascending=False).head(10)
Like

```

```

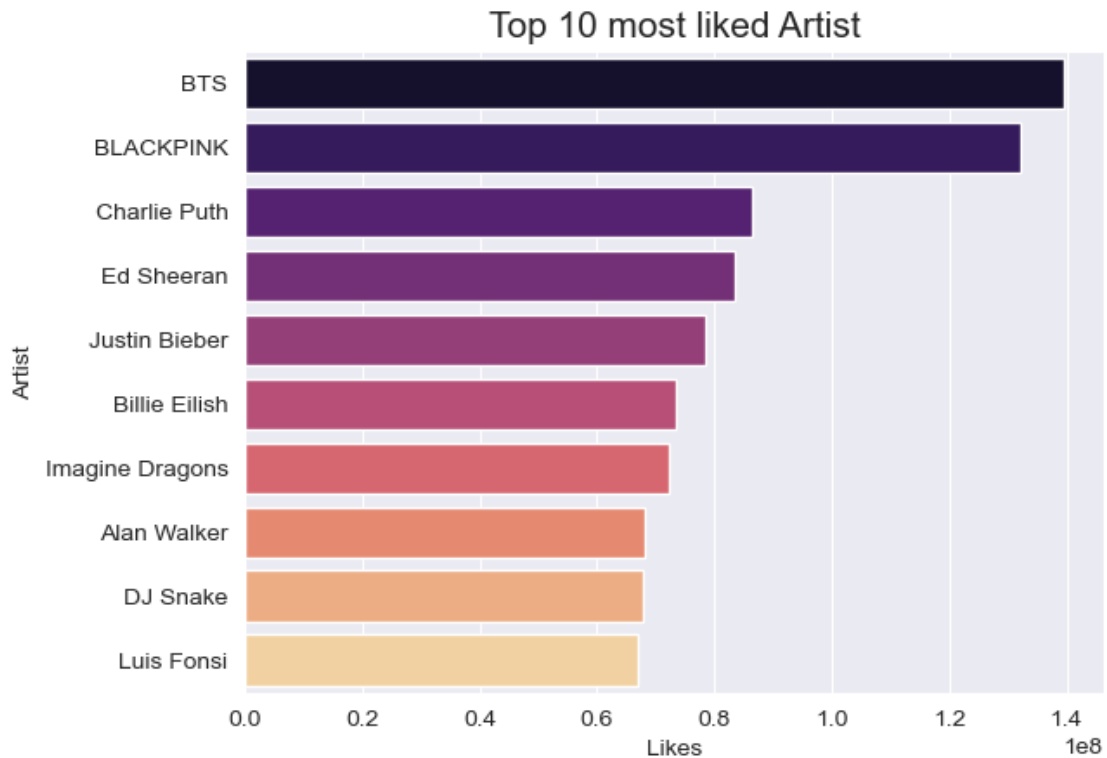
[23]:

```

| | Artist | Likes |
|-----|-----------------|-----------|
| 151 | BTS | 139403808 |
| 149 | BLACKPINK | 132302505 |
| 329 | Charlie Puth | 86336925 |
| 516 | Ed Sheeran | 83543319 |
| 908 | Justin Bieber | 78579227 |
| 199 | Billie Eilish | 73338537 |
| 764 | Imagine Dragons | 72274477 |
| 59 | Alan Walker | 67977702 |

| | | |
|------|------------|----------|
| 411 | DJ Snake | 67775831 |
| 1129 | Luis Fonsi | 66908489 |

```
[24]: #Barplot
sns.set_style('darkgrid')
sns.barplot(data = Like,y='Artist',x='Likes',palette='magma')
plt.title('Top 10 most liked Artist',fontsize=15)
plt.show()
```



```
[25]: #Collecting top 10 Artist whose music video got most comments
comnt = df.groupby('Artist')['Comments'].sum().reset_index().
        sort_values('Comments',ascending=False).head(10)
comnt
```

| | Artist | Comments |
|------|---------------|----------|
| 151 | BTS | 39542659 |
| 149 | BLACKPINK | 19605027 |
| 1722 | Stray Kids | 8458365 |
| 1751 | TWICE | 8384985 |
| 1410 | PSY | 7311608 |
| 908 | Justin Bieber | 7258967 |
| 509 | EXO | 6981950 |

| | | |
|------|---------------|---------|
| 714 | Halsey | 6981661 |
| 1390 | One Direction | 6360892 |
| 1129 | Luis Fonsi | 4703428 |

```
[26]: #Barplot
cmt = px.bar(comnt,x='Artist',y='Comments',color='Artist',title='Top 10 (most_
↳comments from fans) Artists')
cmt.show()
```

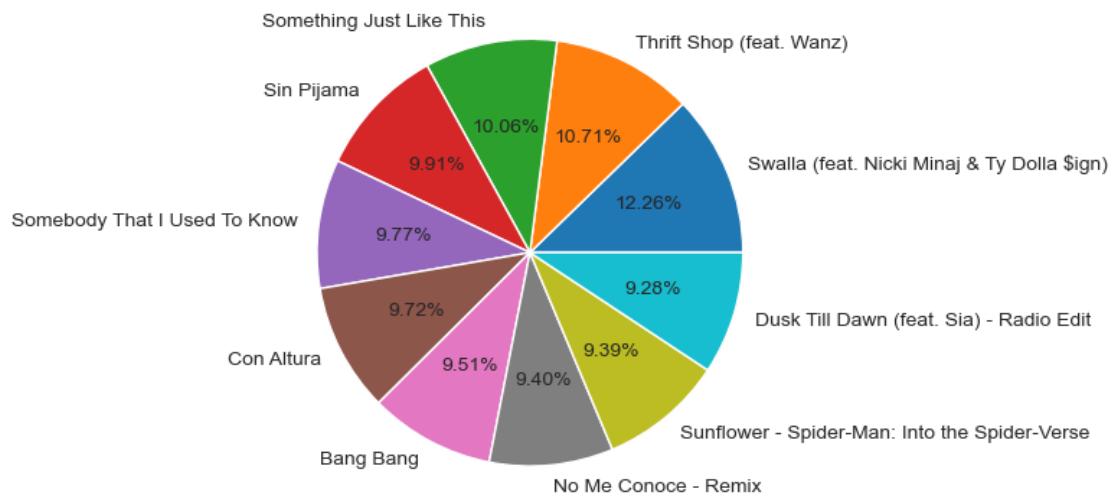
```
[27]: #Collecting top 10 songs which got maximum views,comments and likes
track = df.groupby('Track').agg({'Views':'sum','Likes':'sum','Comments':'sum'}).
↳reset_index().sort_values(by='Views',ascending=False).head(10)
track
```

```
[27]:
```

| | Track | Views | Likes \ |
|-------|---|--------------|----------|
| 13145 | Swalla (feat. Nicki Minaj & Ty Dolla \$ign) | 5.162404e+09 | 28864162 |
| 14152 | Thrift Shop (feat. Wanz) | 4.508467e+09 | 30242165 |
| 12669 | Something Just Like This | 4.236043e+09 | 20565056 |
| 12412 | Sin Pijama | 4.172962e+09 | 17249780 |
| 12643 | Somebody That I Used To Know | 4.114891e+09 | 29168472 |
| 2830 | Con Altura | 4.093098e+09 | 18536081 |
| 1480 | Bang Bang | 4.002769e+09 | 22355292 |
| 9682 | No Me Conoce - Remix | 3.958118e+09 | 11841081 |
| 13071 | Sunflower - Spider-Man: Into the Spider-Verse | 3.954778e+09 | 27499619 |
| 3875 | Dusk Till Dawn (feat. Sia) - Radio Edit | 3.907550e+09 | 28299420 |

| | Comments |
|-------|----------|
| 13145 | 759525 |
| 14152 | 1264245 |
| 12669 | 540890 |
| 12412 | 732738 |
| 12643 | 1586010 |
| 2830 | 562026 |
| 1480 | 578607 |
| 9682 | 215069 |
| 13071 | 662127 |
| 3875 | 793267 |

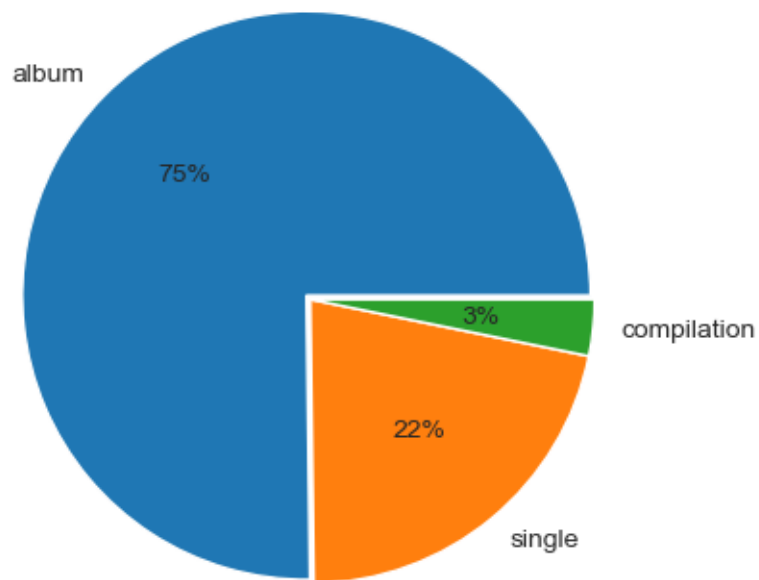
```
[28]: #Pieplot
plt.pie(data=track,x=track.Views,autopct='%.2f%%',labels='Track')
plt.show()
```

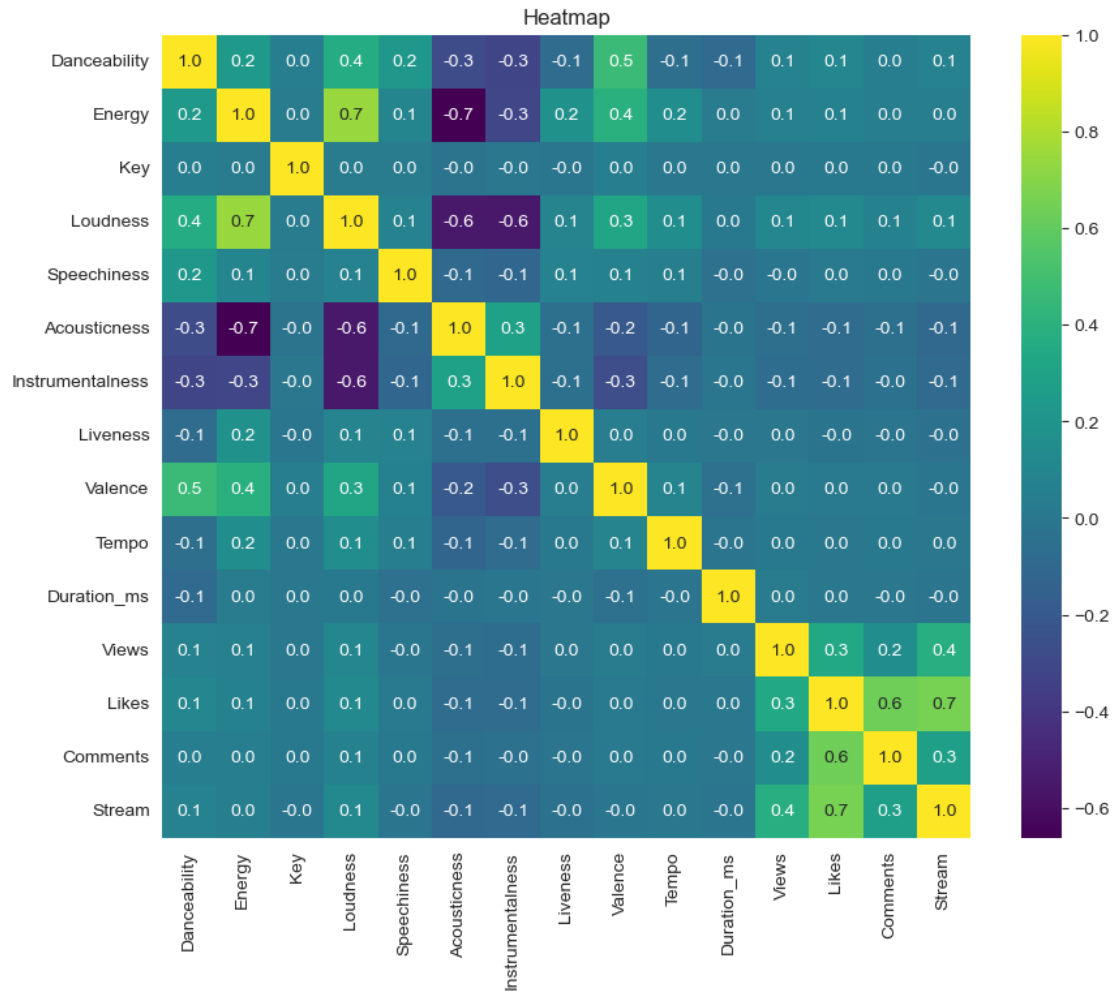
```
[29]: #Types of Album available in data which got most view
Album = df.groupby('Album_type')['Views'].sum().reset_index().
        ↪sort_values('Views',ascending=False)
Album
```

```
[29]:      Album_type      Views
0         album  1.148762e+12
2         single  3.300549e+11
1  compilation  4.882922e+10
```

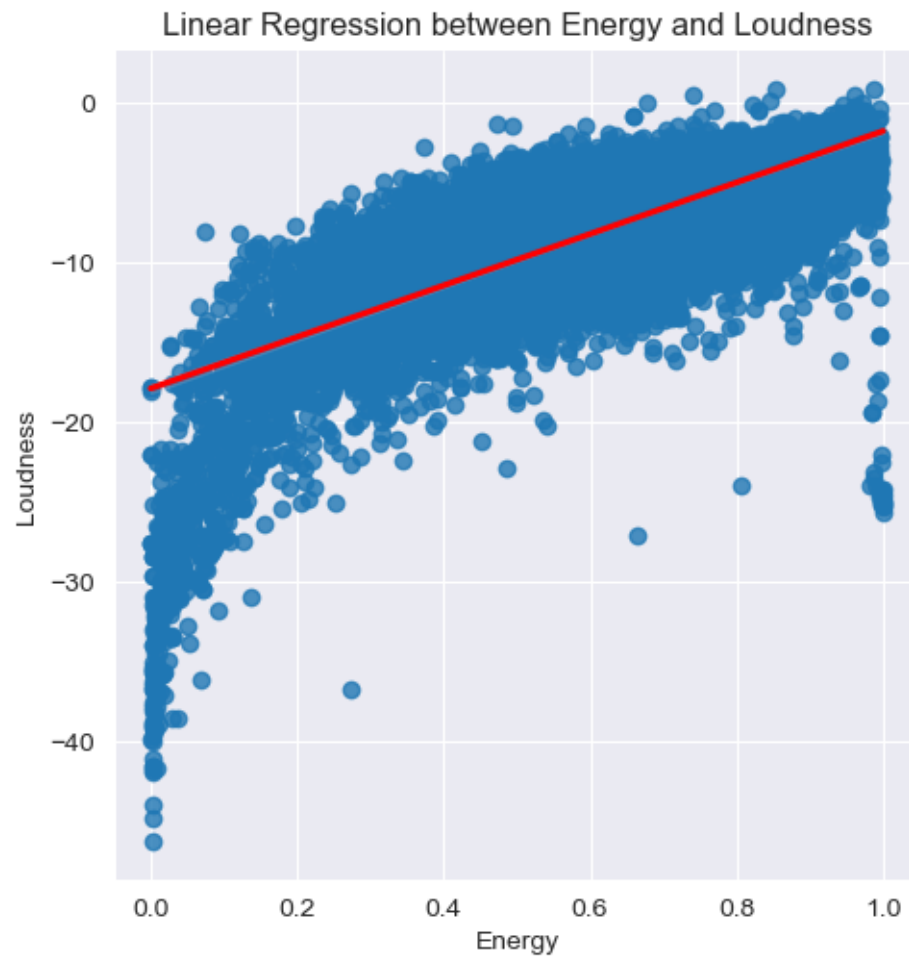
```
[30]: #Visualize through pie chart
plt.pie(data=Album,x='Views',autopct='%f%%',explode=(0.
        ↪02,0,0),labels='Album_type')
plt.show()
```



```
[31]: #The correlation Between our Data's  
plt.figure(figsize=(10,8))  
sns.heatmap(data=df.corr(),annot=True,fmt='.1f',cmap='viridis')  
plt.title('Heatmap')  
plt.show()
```



```
[32]: #Relation between energy and loudness
#using lmplot()
sns.lmplot(data=df,x='Energy',y='Loudness',line_kws={'color':'red'})
plt.title('Linear Regression between Energy and Loudness')
plt.show()
```



[]: