

Dataset Name: Spotify Song Tracks

Dataset

Objectives

Below are the objectives that I've tried to find the answers of doing EDA (Exploratory Data Analysis)

- Top 5 most popular artists
- Top 5 loudest tracks
- Artist with the most danceability song
- Top 10 instrumentalness tracks
- Acousticness of tracks wil popolarity more than 70

Import libraries

```
In [108... import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

Load Dataset

```
In [109... df=pd.read_csv("C:/Users/Admin/Desktop/Programs/Datasets/Spotify/SpotifyFeatures.csv")
df.drop("time_signature", axis=1, inplace=True)
df.head()
```

Out[109]:

	genre	artist_name	track_name	track_id	popularity	acousticness	danceability	duration
0	Movie	Henri Salvador	C'est beau de faire un Show	0BRjO6ga9RKCKjfDqeFgWV	0	0.611	0.389	9
1	Movie	Martin & les fées	Perdu d'avance (par Gad Elmaleh)	0BjC1NfoEOOusryehmNudP	1	0.246	0.590	13
2	Movie	Joseph Williams	Don't Let Me Be Lonely Tonight	0CoSDzoNIKCRs124s9uTVy	3	0.952	0.663	17
3	Movie	Henri Salvador	Dis-moi Monsieur Gordon Cooper	0Gc6TVm52BwZD07Ki6tlvf	0	0.703	0.240	15
4	Movie	Fabien Nataf	Ouverture	0lusIXpMROHdEPvSI1fTQK	4	0.950	0.331	8

Data Cleaning

```
In [110... df.nunique()
```

```
Out[110]: genre                27
artist_name          14564
track_name           148615
track_id             176774
popularity            101
acousticness          4734
danceability          1295
duration_ms           70749
energy                2517
instrumentalness       5400
key                   12
liveness              1732
loudness              27923
mode                  2
speechiness           1641
tempo                 78512
valence               1692
dtype: int64
```

```
In [111... df.isnull().sum()
```

```
Out[111]: genre                0
artist_name          0
track_name           0
track_id             0
popularity            0
acousticness          0
danceability          0
duration_ms           0
energy                0
instrumentalness       0
key                   0
liveness              0
loudness              0
mode                  0
speechiness           0
tempo                 0
valence               0
dtype: int64
```

```
In [112... df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 232725 entries, 0 to 232724
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype
---  -
0   genre                  232725 non-null object
1   artist_name            232725 non-null object
2   track_name             232725 non-null object
3   track_id               232725 non-null object
4   popularity              232725 non-null int64
5   acousticness           232725 non-null float64
6   danceability            232725 non-null float64
7   duration_ms            232725 non-null int64
8   energy                  232725 non-null float64
9   instrumentalness        232725 non-null float64
10  key                     232725 non-null object
11  liveness                232725 non-null float64
12  loudness                232725 non-null float64
13  mode                    232725 non-null object
14  speechiness             232725 non-null float64
15  tempo                   232725 non-null float64
16  valence                  232725 non-null float64
dtypes: float64(9), int64(2), object(6)
memory usage: 30.2+ MB
```

```
In [113]: df.shape
```

```
Out[113]: (232725, 17)
```

```
In [114]: len(df.columns)
```

```
Out[114]: 17
```

```
In [115]: df.describe()
```

```
Out[115]:
```

	popularity	acousticness	danceability	duration_ms	energy	instrumentalness	liv
count	232725.000000	232725.000000	232725.000000	2.327250e+05	232725.000000	232725.000000	232725.0
mean	41.127502	0.368560	0.554364	2.351223e+05	0.570958	0.148301	0.2
std	18.189948	0.354768	0.185608	1.189359e+05	0.263456	0.302768	0.1
min	0.000000	0.000000	0.056900	1.538700e+04	0.000020	0.000000	0.0
25%	29.000000	0.037600	0.435000	1.828570e+05	0.385000	0.000000	0.0
50%	43.000000	0.232000	0.571000	2.204270e+05	0.605000	0.000044	0.1
75%	55.000000	0.722000	0.692000	2.657680e+05	0.787000	0.035800	0.2
max	100.000000	0.996000	0.989000	5.552917e+06	0.999000	0.999000	1.0

```
In [116]: df.corr()
```

Out[116]:

	popularity	acousticness	danceability	duration_ms	energy	instrumentalness	liveness
popularity	1.000000	-0.381295	0.256564	0.002348	0.248922	-0.210983	-0.167995
acousticness	-0.381295	1.000000	-0.364546	0.011203	-0.725576	0.316154	0.069004
danceability	0.256564	-0.364546	1.000000	-0.125781	0.325807	-0.364941	-0.041684
duration_ms	0.002348	0.011203	-0.125781	1.000000	-0.030550	0.076021	0.023783
energy	0.248922	-0.725576	0.325807	-0.030550	1.000000	-0.378957	0.192801
instrumentalness	-0.210983	0.316154	-0.364941	0.076021	-0.378957	1.000000	-0.134198
liveness	-0.167995	0.069004	-0.041684	0.023783	0.192801	-0.134198	1.000000
loudness	0.363011	-0.690202	0.438668	-0.047618	0.816088	-0.506320	0.045686
speechiness	-0.151076	0.150935	0.134560	-0.016171	0.145120	-0.177147	0.510147
tempo	0.081039	-0.238247	0.021939	-0.028456	0.228774	-0.104133	-0.051355
valence	0.060076	-0.325798	0.547154	-0.141811	0.436771	-0.307522	0.011804

Data Analysis

Top 5 most popular artists

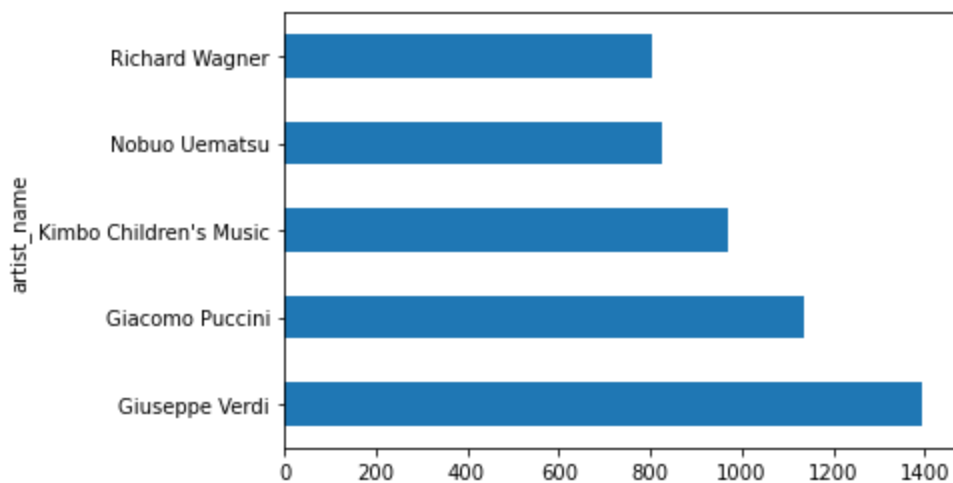
In [117... `top_five_artists = df.groupby("artist_name").count().sort_values(by="track_name", ascend`
`top_five_artists`

Out[117]:

artist_name	
Giuseppe Verdi	1394
Giacomo Puccini	1137
Kimbo Children's Music	971
Nobuo Uematsu	825
Richard Wagner	804

Name: track_name, dtype: int64

In [118... `top_five_artists.plot.barh()`
`plt.show()`



Top 5 loudest tracks

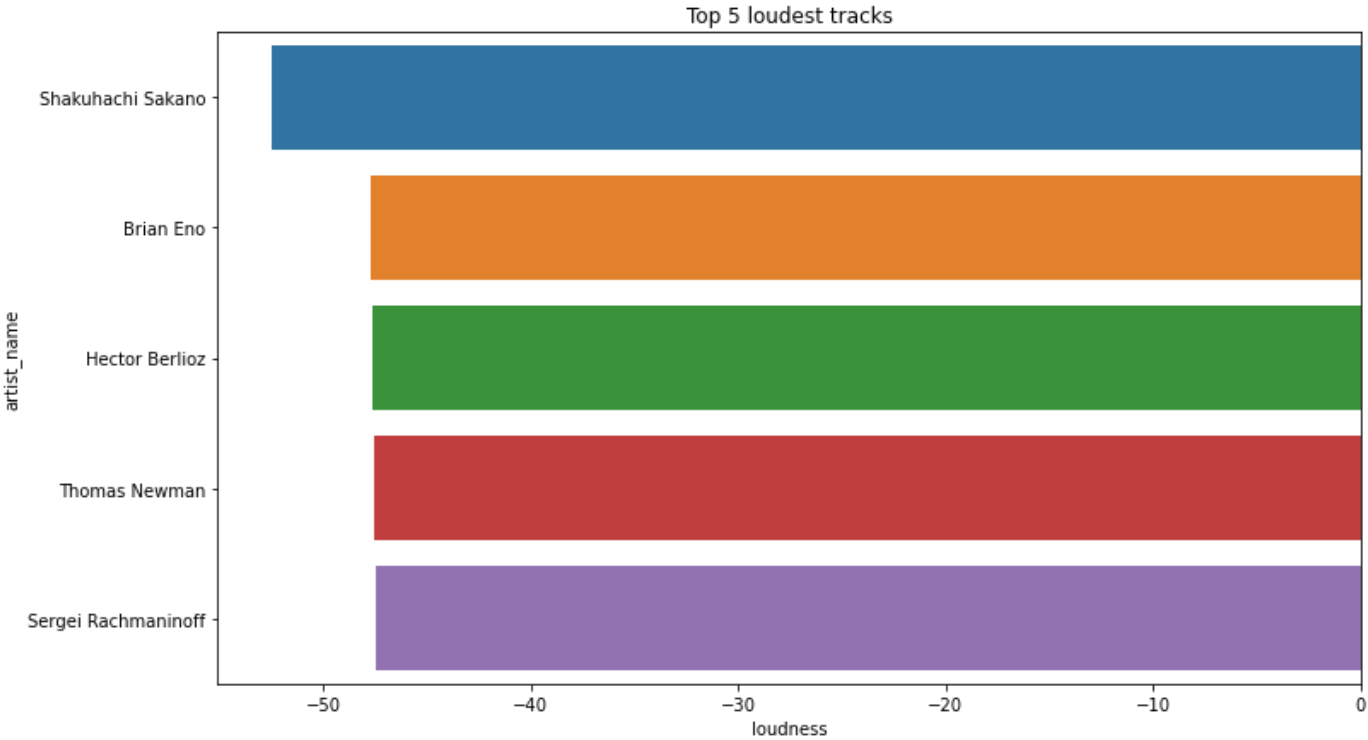
In [119... `top_five_loudest_track=df[["loudness","track_name","artist_name"]].sort_values(by="loudn`
`top_five_loudest_track`

Out[119]:

	loudness	track_name	artist_name
218717	-52.457	Call to Wake	Shakuhachi Sakano
218540	-47.669	Neroli	Brian Eno
126046	-47.599	La Damnation de Faust, Op. 24, H. 111: Pt. II,...	Hector Berlioz
203063	-47.499	Limes - Instrumental	Thomas Newman
126602	-47.432	Variations On A Theme Of Chopin, Op.22: Variat...	Sergei Rachmaninoff

In [120...

```
plt.figure(figsize=(12,7))
sns.barplot(x="loudness", y="artist_name", data=top_five_loudest_track)
plt.title("Top 5 loudest tracks")
plt.show()
```



Artist with the most danceability song

In [121...

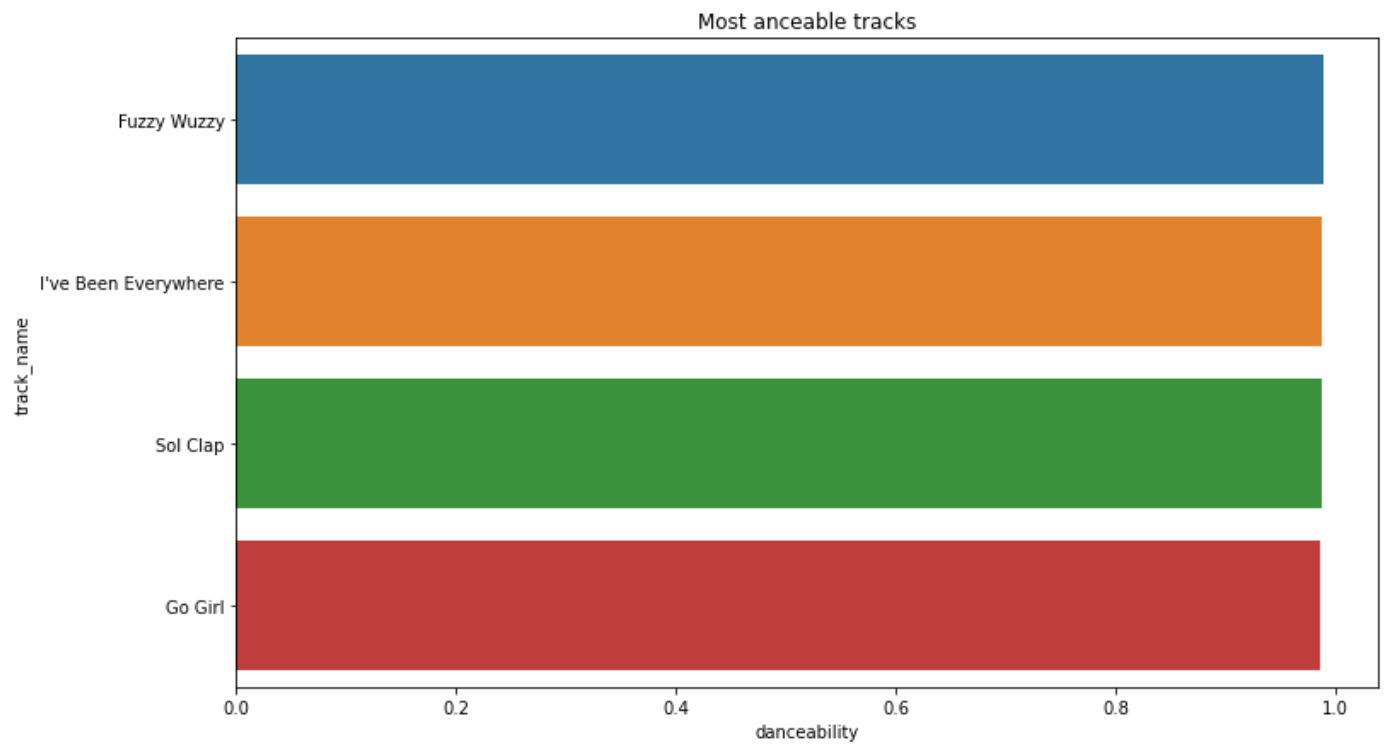
```
most_danceable_track=df[["danceability","track_name","artist_name"]].sort_values(by="dan
most_danceable_track
```

Out[121]:

	danceability	track_name	artist_name
75396	0.989	Fuzzy Wuzzy	Juice Music
75762	0.987	I've Been Everywhere	Juice Music
26911	0.987	Sol Clap	Quantic
178675	0.987	Sol Clap	Quantic
90205	0.986	Go Girl	Pitbull

In [122...

```
plt.figure(figsize=(12,7))
sns.barplot(x="danceability",y="track_name",data=most_danceable_track)
plt.title("Most anceable tracks")
plt.show()
```



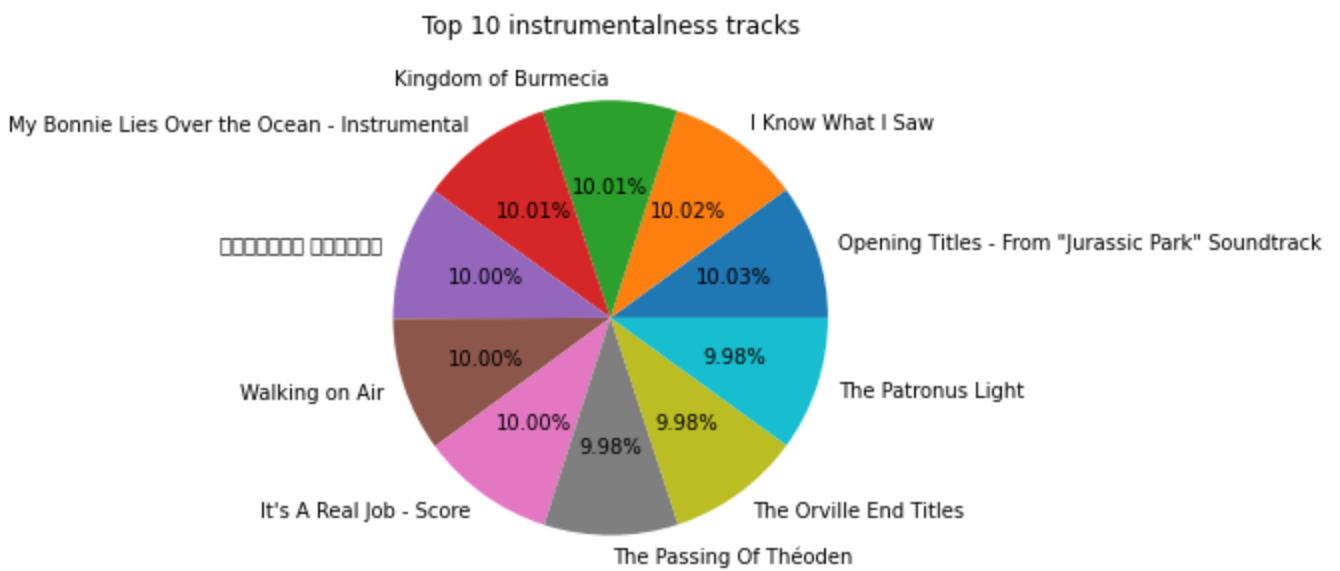
Top 10 instrumentality tracks

```
In [123]: top_ten_instrumentality_track=df[["instrumentality","track_name","artist_name"]].sort_
top_ten_instrumentality_track
```

Out[123]:

	instrumentality	track_name	artist_name
199660	0.999	Opening Titles - From "Jurassic Park" Soundtrack	John Williams
203562	0.998	I Know What I Saw	Kyle Dixon & Michael Stein
31144	0.997	Kingdom of Burmecia	Nobuo Uematsu
76749	0.997	My Bonnie Lies Over the Ocean - Instrumental	Children Songs Company
34781	0.996	スリルのテーマ ～サスペンス	Capcom Sound Team
218722	0.996	Walking on Air	Shakuhachi Sakano
205376	0.996	It's A Real Job - Score	John Debney
199672	0.994	The Passing Of Théoden	Howard Shore
194160	0.994	The Orville End Titles	Bruce Broughton
198336	0.994	The Patronus Light	John Williams

```
In [128]: plt.figure(figsize=(12,5))
plt.pie(x="instrumentality",data=top_ten_instrumentality_track,autopct='%1.2f%%',label
plt.title("Top 10 instrumentality tracks")
plt.show()
```



Acousticness of tracks wil popularity more than 70

```
In [125]: popularity_more_than_70=df[df.popularity > 70].sort_values(by="popularity", ascending=False)
popularity_more_than_70
```

```
Out[125]:
```

	genre	artist_name	track_name	track_id	popularity	acousticness	danceability
9027	Dance	Ariana Grande	7 rings	14msK75pk3pA33pzPVNtBF	100	0.5780	0.725
107804	Pop	Ariana Grande	7 rings	14msK75pk3pA33pzPVNtBF	100	0.5780	0.725
86951	Rap	Post Malone	Wow.	6MWtB6iiXylwun0YzU6DFP	99	0.1630	0.833
107802	Pop	Ariana Grande	break up with your girlfriend, i'm bored	4kV4N9D1iKVxx1KLvtTpjS	99	0.0421	0.726
107803	Pop	Post Malone	Wow.	6MWtB6iiXylwun0YzU6DFP	99	0.1630	0.833
...
110417	Pop	Jason Derulo	It Girl	4fINc8dnfcz7AdhFYVA4i7	71	0.0165	0.668
110407	Pop	Bright Eyes	First Day Of My Life	5OiaAaIMYICZONyDBxqk4G	71	0.9150	0.468
110396	Pop	BTS	Airplane pt.2	1trFxVLL8WKhYap543e74I	71	0.0268	0.769
110387	Pop	Dr. Dre	The Next Episode	4LwU4Vp6od3Sb08CsP99GC	71	0.0279	0.916
226424	Rock	La Mosca Tse-Tse	Para No Verte Más	19CmuECYssqkPWANF4nLWM	71	0.0168	0.658

7559 rows × 17 columns

```
In [134]: sns.displot(popularity_more_than_70["acousticness"],color="green")
plt.title('Acoustiness for Songs with More than 70 Popularity')
plt.show()
```

Acoustiness for Songs with More than 70 Popularity

