MUSICAL SONGS ANALYSIS

A data analysis and visualization project

Spotify is a Swedish audio streaming and media services provider founded in April 2006. It is the world's largest music streaming service provider and has over 381 million monthly active users, which also includes172 million paid subscribers. We will be exploring and qualify data about music and drawing valuable insights

This dataset contains information regarding song's genre, artists name, tracks name, songs popularity, acousticness, danceability, loudness, artist name and id.

Project Scope

In this project we are going to find the most listening songs, most rated songs, popularity of songs, year wise songs distribution, correlation between variables, remove empty values, merge two csv files, different graphs and plots to show comparison and relation between various fields and columns of dataset, duration and can predict the upcoming songs popularity and success ration.

Link to Dataset

https://www.kaggle.com/datasets/zaheenhamidani/ultimate-spotify-tracks-db

https://www.kaggle.com/datasets/lehaknarnauli/spotify-datasets?select=artists.csv

Queries

1.find the top 5 most popular songs

2. find out artist with most danceability songs

3. find the most instrumentalness songs

4.find out duration of songs in different genres.

5.find out the total no of songs per year

6.transform the release date(dd/mm/yy) as showing only year(yy)

7.find correlation heatmap between any two variables

8.convert duration from milliseconds to seconds only

9.find out the relation between loudness and energy

10. Creating Visualization with Correlation Using Pearson Method

Importing libraries

```
import numpy as np
from sklearn import datasets
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
```

Reading csv file1

```
a1=pd.read_csv('C:/Users/pc/OneDrive/Desktop/DAV/tracks.csv')
a1.head()
```

	id	name	popularity	duration_ms	explicit	artists	id
0	35iwgR4jXetI318WEWsa1Q	Carve	6	126903	0	['Uli']	['45tIt06XoI0Iio4L
1	021ht4sdgPcrDgSk7JTbKY	Capítulo 2.16 - Banquero Anarquista	0	98200	0	['Fernando Pessoa']	['14jtPCOoNZwquk5wc
2	07A5yehtSnoedViJAZkNnc	Vivo para Quererte - Remasterizado	0	181640	0	['Ignacio Corsini']	['5LiOoJbxVSAMkBS2fl
3	08FmqUhxtyLTn6pAh6bk45	El Prisionero - Remasterizado	0	176907	0	['Ignacio Corsini']	['5LiOoJbxVSAMkBS2fl
4	08y9GfoqCWfOGsKdwojr5e	Lady of the Evening	0	163080	0	['Dick Haymes']	['3BiJGZsyX9sJchTqc

Reading csv file2

```
a2=pd.read_csv('C:/Users/pc/OneDrive/Desktop/DAV/SpotifyFeatures.csv')
a2.head()
```

	genre	artist_name	track_name	track_id	popularity	acousticness	danceability	durati
0	Movie	Henri Salvador	C'est beau de faire un Show	0BRjO6ga9RKCKjfDqeFgWV	0	0.611	0.389	
1	Movie	Martin & les fées	Perdu d'avance (par Gad Elmaleh)	0BjC1NfoEOOusryehmNudP	1	0.246	0.590	1
2	Movie	Joseph Williams	Don't Let Me Be Lonely Tonight	0CoSDzoNIKCRs124s9uTVy	3	0.952	0.663	1
3	Movie	Henri Salvador	Dis-moi Monsieur Gordon Cooper	0Gc6TVm52BwZD07Ki6tIvf	0	0.703	0.240	1

Finding null values

```
pd.isnull(a1).sum()
```

```
id                  0
name               71
popularity          0
duration_ms         0
explicit            0
artists             0
id_artists          0
release_date        0
danceability        0
energy              0
key                 0
loudness            0
mode                0
speechiness         0
acousticness        0
instrumentalness    0
liveness            0
valence             0
tempo               0
```

Counting total null values

```
pd.isnull(a1).sum().sum()
```

```
71
```

Information about database

```
a1.info
```

```
<bound method DataFrame.info of                                  id
\
0       35iwgR4jXetI318WEWsa1Q                                Carve
1       021ht4sdgPcrDgSk7JTbKY  Capítulo 2.16 - Banquero Anarquista
2       07A5yehtSnoedViJAZkNnc   Vivo para Quererte - Remasterizado
3       08FmqUhxtyLTn6pAh6bk45        El Prisionero - Remasterizado
4       08y9GfoqCWfOGsKdwojr5e                  Lady of the Evening
...                        ...                                 ...
586667  5rgu12WBIHQtvej2MdHSH0                                云与海
586668  0NuWgxEp51CutD2pJoF4OM                               blind
586669  27Y1N4Q4U3EfDU5Ubw8ws2           What They'll Say About Us
586670  45XJsGpFTyzbzeWK8VzR8S                     A Day At A Time
586671  5Ocn6dZ3BJFPWh4ylwFXtn                    Mar de Emociones

        popularity  duration_ms  explicit                        artists  \
0                6       126903         0                       ['Uli']
1                0        98200         0            ['Fernando Pessoa']
2                0       181640         0            ['Ignacio Corsini']
3                0       176907         0            ['Ignacio Corsini']
```

Sorting according to release date by this year to last years

```
sort_a1=a1.sort_values('release_date', ascending=False).head(8)
sort_a1
```

	id	name	popularity	duration_ms	explicit	artists	
93919	6cjSDzbazK0wsWWbZMRxsl	Blow Your Mind (Mwah)	0	178583	0	['Dua Lipa']	['6M2wZ9GZgrQXH
188952	6b3AippDKmNU9ZpYv1u9L4	Rojo	0	151147	0	['J Balvin']	['1vyhD5VmyZ7KMfV
188868	1YTswXgjPiwAIQIMlCWsZV	Negro	0	182507	0	['J Balvin']	['1vyhD5VmyZ7KMfV
188867	1WEMrMUauyF2FgoOfyAIDb	Machika	0	189653	0	['J Balvin', 'Jeon', 'Anitta']	['1vyhD5VmyZ7KMfV '3ECsaSd1fon
188866	1T01S75AMk2nXp2ExJA8tl	Negro	0	182507	0	['J Balvin']	['1vyhD5VmyZ7KMfV
188865	1OZgkgzRuKafaGXlrtVFVh	Morado	0	200667	0	['J Balvin']	['1vyhD5VmyZ7KMfV
188864	1OOUyI2vvPGHfiLYvlASnm	LA CANCIÓN	0	242573	0	['J Balvin', 'Bad Bunny']	['1vyhD5VmyZ7KMfV '4q3ewBCX7sl

Descriptive statistics of Spotify tracks

```
a1.describe().transpose()
```

	count	mean	std	min	25%	50%	75%
popularity	586672.0	27.570053	18.370642	0.0	13.0000	27.000000	41.00000
duration_ms	586672.0	230051.167286	126526.087418	3344.0	175093.0000	214893.000000	263867.00000
explicit	586672.0	0.044086	0.205286	0.0	0.0000	0.000000	0.00000
danceability	586672.0	0.563594	0.166103	0.0	0.4530	0.577000	0.68600
energy	586672.0	0.542036	0.251923	0.0	0.3430	0.549000	0.74800
key	586672.0	5.221603	3.519423	0.0	2.0000	5.000000	8.00000
loudness	586672.0	-10.206067	5.089328	-60.0	-12.8910	-9.243000	-6.48200
mode	586672.0	0.658797	0.474114	0.0	0.0000	1.000000	1.00000
speechiness	586672.0	0.104864	0.179893	0.0	0.0340	0.044300	0.07630
acousticness	586672.0	0.449863	0.348837	0.0	0.0969	0.422000	0.78500
instrumentalness	586672.0	0.113451	0.266868	0.0	0.0000	0.000024	0.00955
liveness	586672.0	0.213935	0.184326	0.0	0.0983	0.139000	0.27800
valence	586672.0	0.552292	0.257671	0.0	0.3460	0.564000	0.76900
tempo	586672.0	118.464857	29.764108	0.0	95.6000	117.384000	136.32100

Top 10 most popular songs

```
most_popular=a1.query('popularity>90',inplace=False).sort_values('popularity',
ascending=False)
most_popular[:10]
```

93802	4iJyoBOLtHqaGxP12qzhQI	Peaches (feat. Daniel Caesar & Giveon)	100	198082	1	['Justin Bieber', 'Daniel Caesar', 'Giveon']	['1uNFoZAHBGt '20wkVLut
93803	7lPN2DXiMsVn7XUKtOW1CS	drivers license	99	242014	1	['Olivia Rodrigo']	['1McMsnEElThX
93804	3Ofmpyhv5UAQ70mENzB277	Astronaut In The Ocean	98	132780	0	['Masked Wolf']	['1uU7g3DNSbsu
92810	5QO79kh1waicV47BqGRL3g	Save Your Tears	97	215627	1	['The Weeknd']	['1Xyo4u8uXC1Zr
92811	6tDDoYIxWvMLTdKpjFkc1B	telepatía	97	160191	0	['Kali Uchis']	['1U1el3k54VvEl
92813	0VjIjW4GlUZAMYd2vXMi3b	Blinding Lights	96	200040	0	['The Weeknd']	['1Xyo4u8uXC1Zr
93805	7MAibcTli4IisCtbHKrGMh	Leave The Door Open	96	242096	0	['Bruno Mars', 'Anderson .Paak', 'Silk	['0du5cEVh5yTK' '3jK9MiCrA

Indexing according to release date

```
a1.set_index("release_date",inplace=True)
a1.index=pd.to_datetime(a1.index)
a1.head(8)
```

release_date	id	name	popularity	duration_ms	explicit	artists	
1922-02-22	35iwgR4jXetI318WEWsa1Q	Carve	6	126903	0	['Uli']	['45tIt
1922-06-01	021ht4sdgPcrDgSk7JTbKY	Capítulo 2.16 - Banquero Anarquista	0	98200	0	['Fernando Pessoa']	['14jtPCOo
1922-03-21	07A5yehtSnoedViJAZkNnc	Vivo para Quererte - Remasterizado	0	181640	0	['Ignacio Corsini']	['5LiOoJbx\
1922-03-21	08FmqUhxtyLTn6pAh6bk45	El Prisionero - Remasterizado	0	176907	0	['Ignacio Corsini']	['5LiOoJbx\
1922-01-01	08y9GfoqCWfOGsKdwojr5e	Lady of the Evening	0	163080	0	['Dick Haymes']	['3BiJGZ

Finding artist at given index

```
a1[["artists"]].iloc[15]
```

```
...    artists     ['Fernando Pessoa']
       Name: 1922-06-01 00:00:00, dtype: object
```

Changing duration from milliseconds to seconds

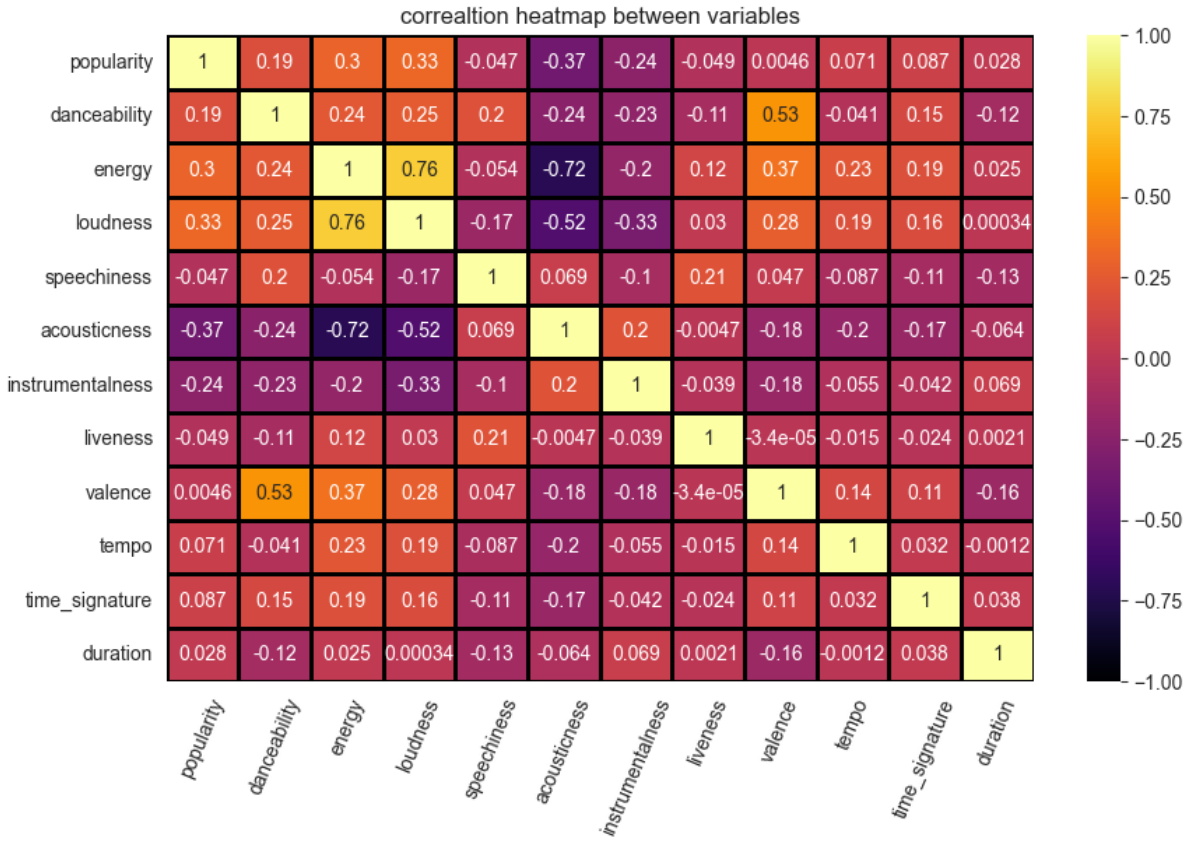
```
a1["duration"]=a1["duration_ms"].apply(lambda x:round(x/1000))
a1.drop("duration_ms",inplace=True,axis=1)
a1.duration.head()
```

```
...    release_date
       1922-02-22    127
       1922-06-01     98
       1922-03-21    182
       1922-03-21    177
       1922-01-01    163
       Name: duration, dtype: int64
```

Correlation heatmap between variables using Pearson method

```
corr_df=a1.drop(["key","mode","explicit"],axis=1).corr(method="pearson")
plt.figure(figsize=(14,6))
heatmap=sns.heatmap(corr_df,annot=Truevmin=-
1,vmax=1,center=0,cmap="inferno",linewidths=1,linecolor="black")
heatmap.set_title("correlation heatmap between variables")
heatmap.set_xticklabels(heatmap.get_xticklabels(),rotation=67)
```

```
[Text(0.5, 0, 'popularity'),
 Text(1.5, 0, 'danceability')
 Text(2.5, 0, 'energy'),
 Text(3.5, 0, 'loudness'),
 Text(4.5, 0, 'speechiness'),
 Text(5.5, 0, 'acousticness')
 Text(6.5, 0, 'instrumentalne
 Text(7.5, 0, 'liveness'),
 Text(8.5, 0, 'valence'),
 Text(9.5, 0, 'tempo'),
 Text(10.5, 0, 'time_signatur
 Text(11.5, 0, 'duration')]
```



Making sample

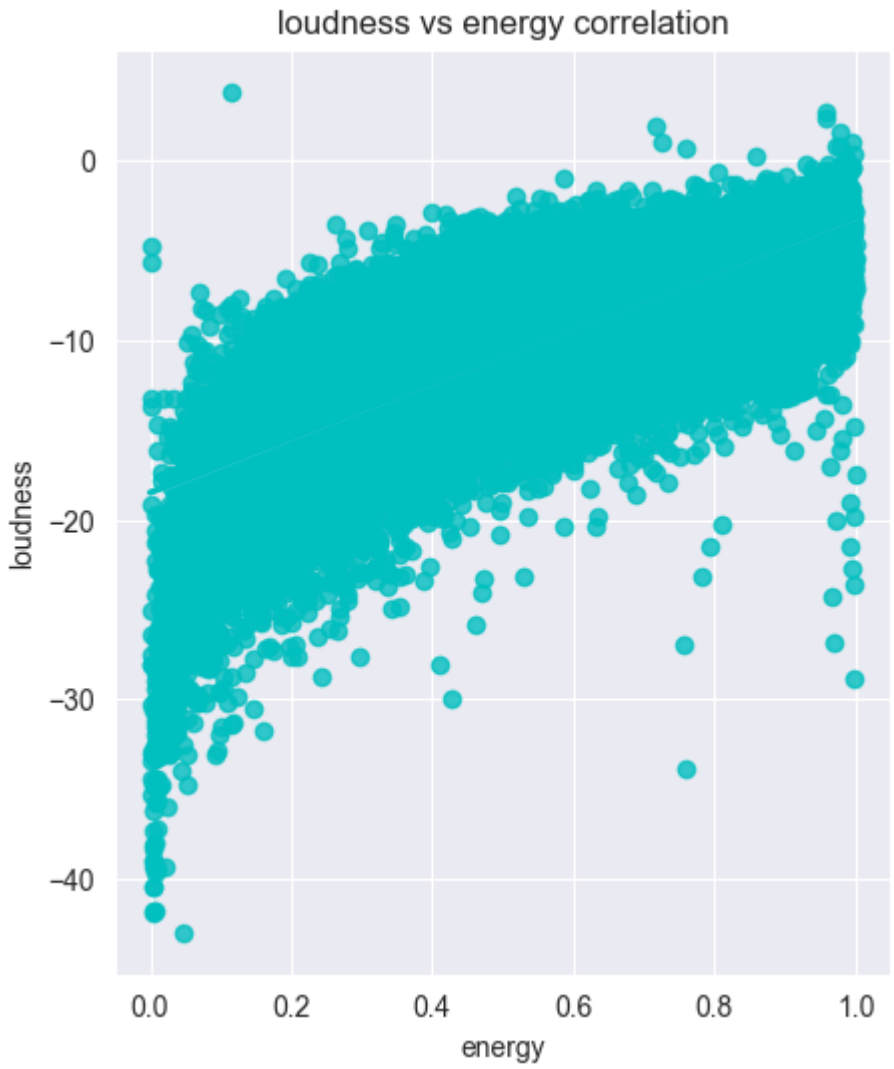
```
sample=a1.sample(int(0.005*len(a1)))
print(sample)
print(len(sample))
```

```
                                id  \
release_date
2015-08-14    7jDKrOCSA26DfGy3ZAVVoI
2014-05-12    5ZVqZ4Aj6ykx3TBaqwj9gV
1928-01-01    3UWvFRdhjsH4656C8d392a
1968-01-01    1a74ta3QcS8sTxnbekVMID
1996-01-24    5aL4lVMBzqrgvu9wCCSJai
...                              ...
2015-11-20    7qPONKgpTIiXM6mDnyhkY1
1961-01-01    4cRJvAh0BwNng8nO5Vnf5F
1995-04-04    2Di9dyLnotlGyKPKCRgKVh
1955-01-01    04zEcp8dG9ixUTG4s3ECaN
1974-11-23    6A41D4U2lpxJs8rPHOxZBk

                                                  name  popularity  \
release_date
2015-08-14    Main Hoon Hero Tera (Armaan Malik Version)          45
2014-05-12                                  Kır Papatyası          45
```

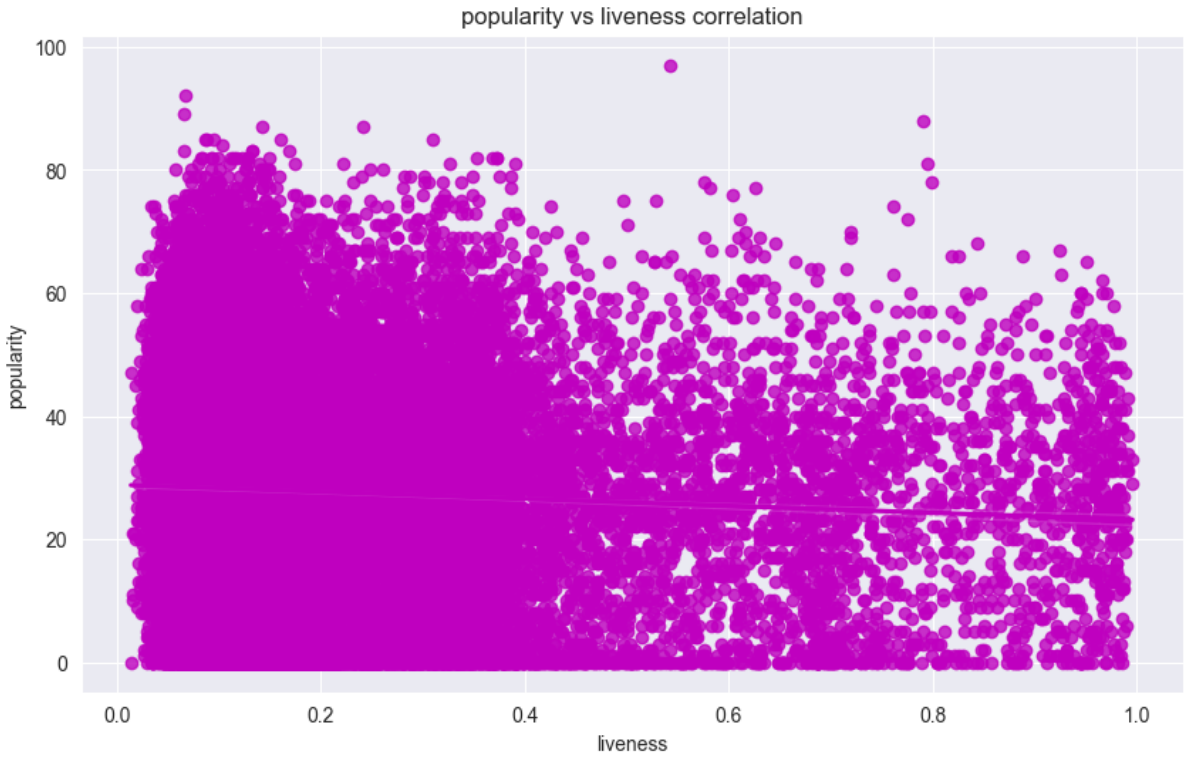
Correlation between loudness and energy

```
plt.figure(figsize=(10,6))
sns.regplot(data=sample,y="loudness",x="energy",color="c").set(title="loudness vs energy correlation")
```



Relation between popularity and correlation

```
plt.figure(figsize=(10,6))
sns.regplot(data=sample,y="popularity",x="liveness",color="m").set(title="popu
larity vs liveness correlation")
```



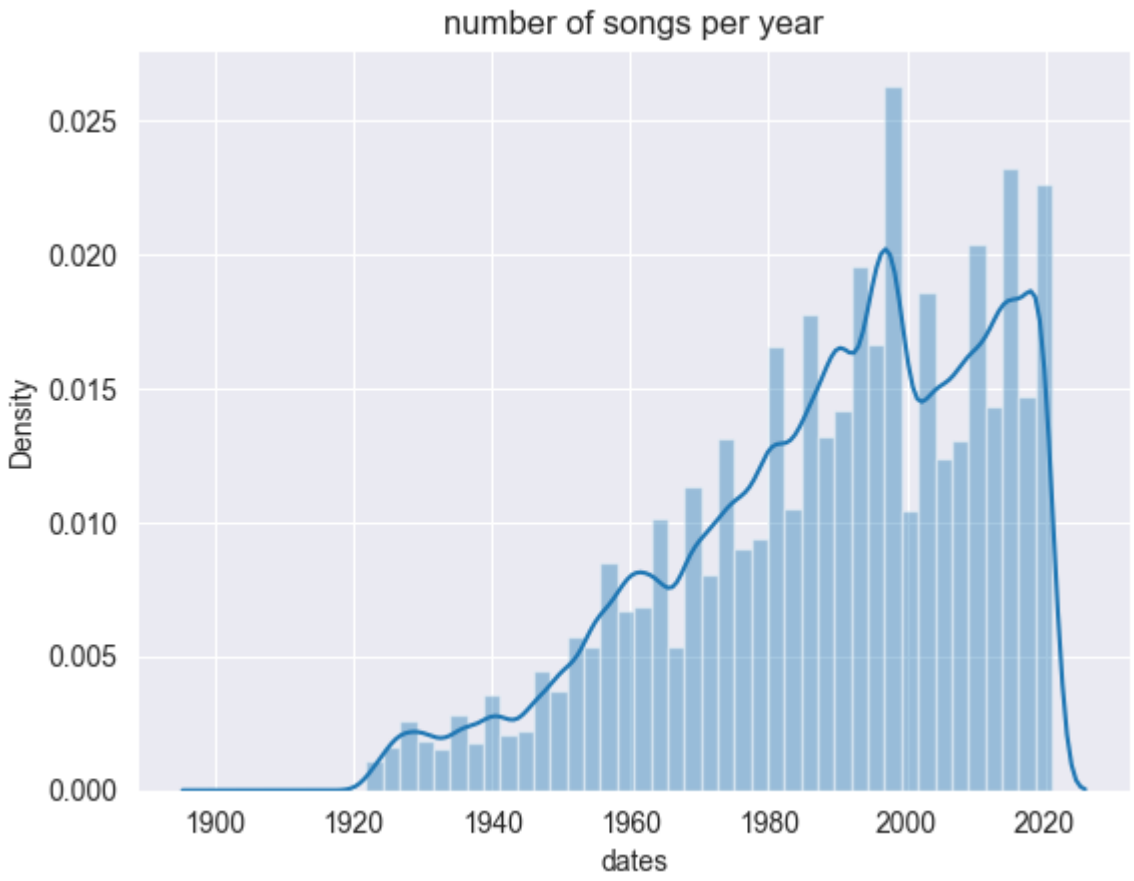
transform the release date(dd/mm/yy) as showing only year(yy)

```
a1['dates']=a1.index.get_level_values('release_date')
a1.dates=pd.to_datetime(a1.dates)
years=a1.dates.dt.year
years
```

```
release_date
1922-02-22    1922
1922-06-01    1922
1922-03-21    1922
1922-03-21    1922
1922-01-01    1922
              ...
2020-09-26    2020
2020-10-21    2020
2020-09-02    2020
2021-03-05    2021
2015-07-01    2015
Name: dates, Length: 586672, dtype: int64
```

No of songs per year

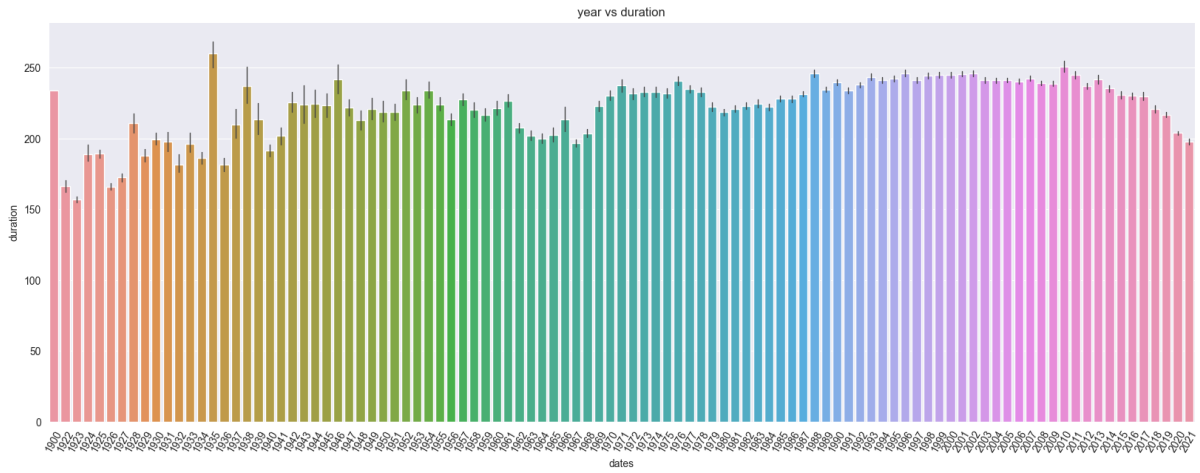
```
sns.distplot(years).set(title="number of songs per year")
```



Year vs songs duration

```
total_dur=a1.duration
fig_dims=(20,7)
fig,ax=plt.subplots(figsize=fig_dims)
fig=sns.barplot(x=years,y=total_dur,ax=ax,errwidth=True).set(title="year vs
duration")
plt.xticks(rotation=60)
```

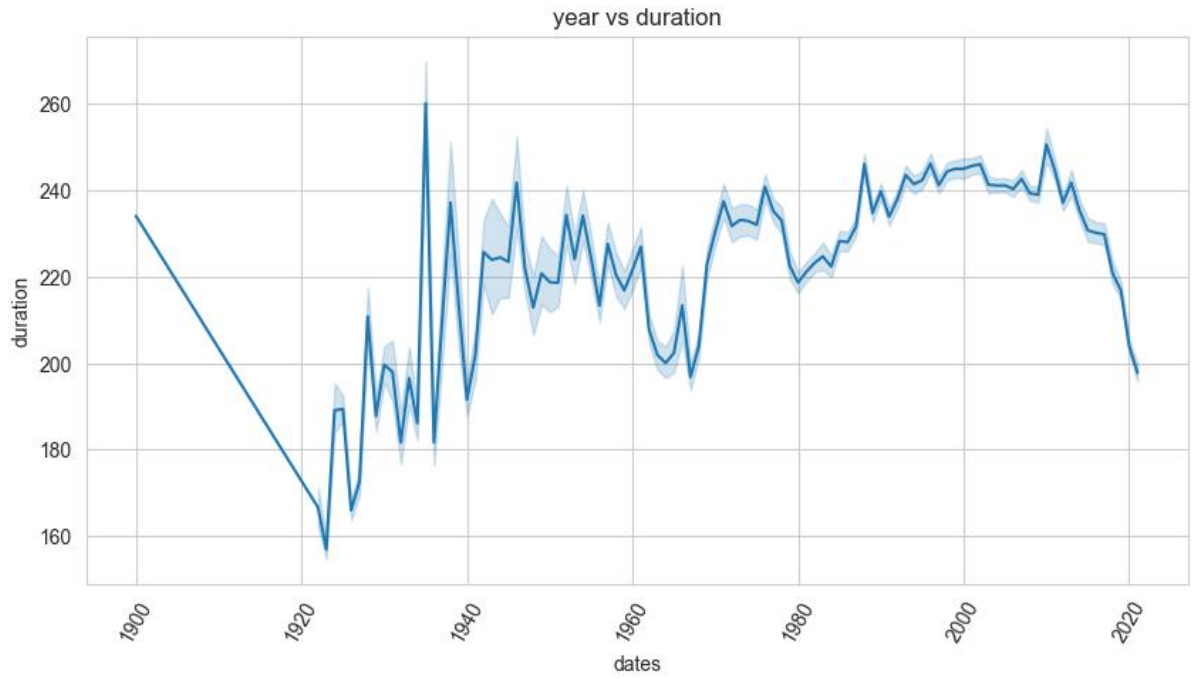
```
(array([  0,   1,   2,   3,   4,   5,   6,   7,   8,   9,  10,  11,  12,
         13,  14,  15,  16,  17,  18,  19,  20,  21,  22,  23,  24,  25,
         26,  27,  28,  29,  30,  31,  32,  33,  34,  35,  36,  37,  38,
         39,  40,  41,  42,  43,  44,  45,  46,  47,  48,  49,  50,  51,
         52,  53,  54,  55,  56,  57,  58,  59,  60,  61,  62,  63,  64,
         65,  66,  67,  68,  69,  70,  71,  72,  73,  74,  75,  76,  77,
         78,  79,  80,  81,  82,  83,  84,  85,  86,  87,  88,  89,  90,
         91,  92,  93,  94,  95,  96,  97,  98,  99, 100]),
 [Text(0, 0, '1900'),
  Text(1, 0, '1922'),
  Text(2, 0, '1923'),
  Text(3, 0, '1924'),
  Text(4, 0, '1925'),
  Text(5, 0, '1926'),
  Text(6, 0, '1927'),
  Text(7, 0, '1928'),
  Text(8, 0, '1929'),
```



Line graph between year and duration

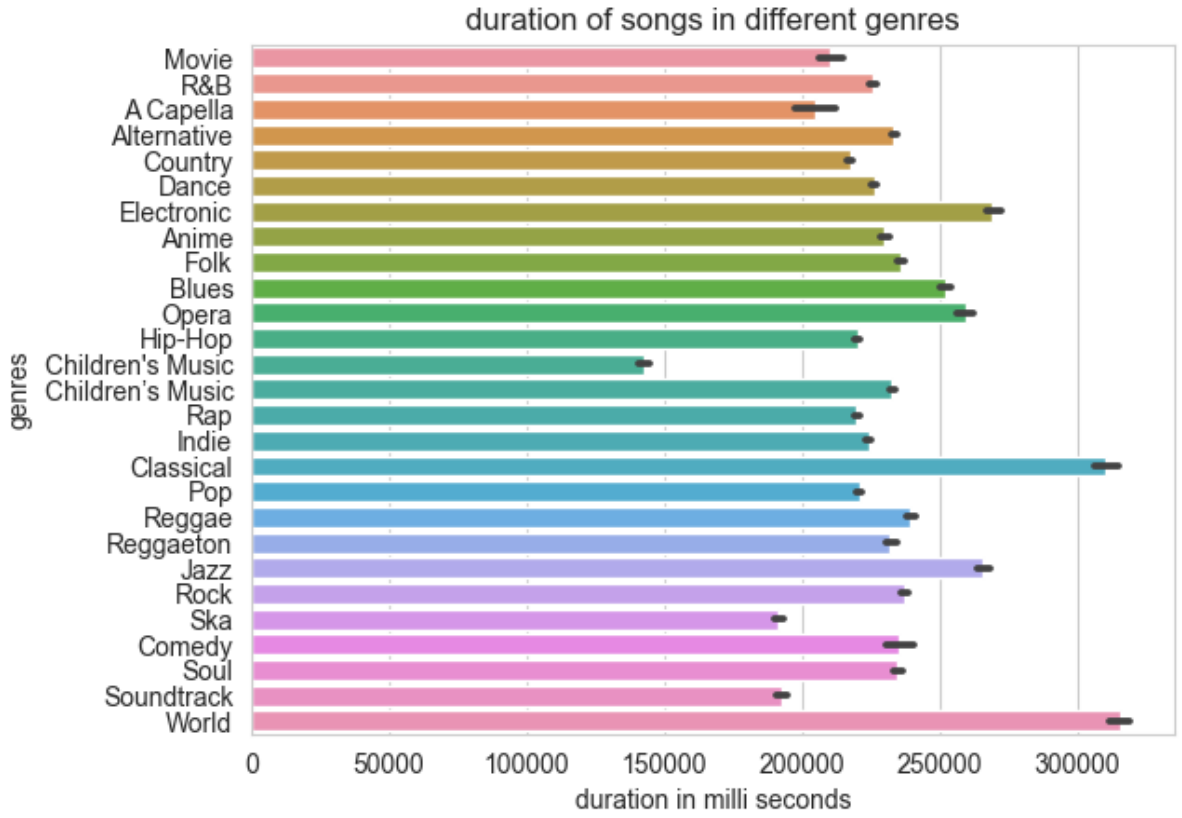
```
total_dur=a1.duration
sns.set_style(style="whitegrid")
fig_dims=(10,5)
fig,ax=plt.subplots(figsize=fig_dims)
fig=sns.lineplot(x=years,y=total_dur,ax=ax).set(title="year vs duration")
plt.xticks(rotation=60)
```

```
(array([1880., 1900., 1920., 1940., 1960., 1980., 2000., 2020., 2040.]),
 [Text(1880.0, 0, '1880'),
  Text(1900.0, 0, '1900'),
  Text(1920.0, 0, '1920'),
  Text(1940.0, 0, '1940'),
  Text(1960.0, 0, '1960'),
  Text(1980.0, 0, '1980'),
  Text(2000.0, 0, '2000'),
  Text(2020.0, 0, '2020'),
  Text(2040.0, 0, '2040')])
```

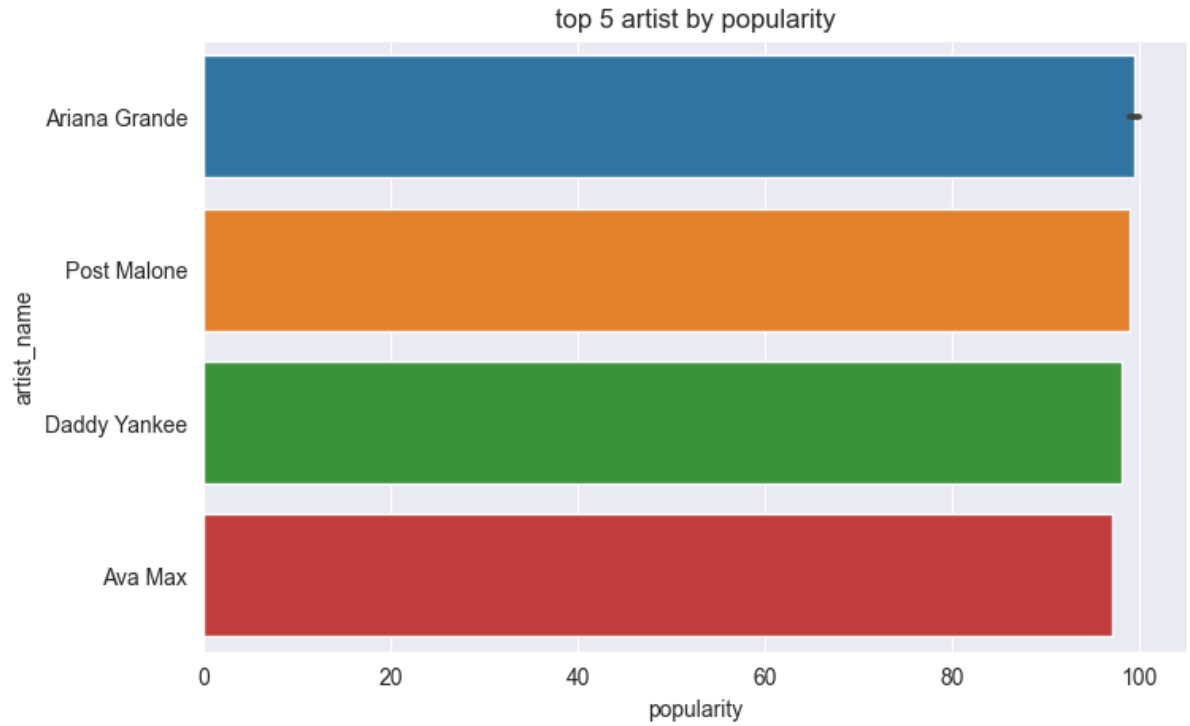



```
plt.title("duration of songs in different genres")
sns.color_palette("rocket",as_cmap=True)
sns.barplot(y="genre",x="duration_ms",data=a2)
plt.xlabel("duration in milli seconds")
plt.ylabel("genres")
```

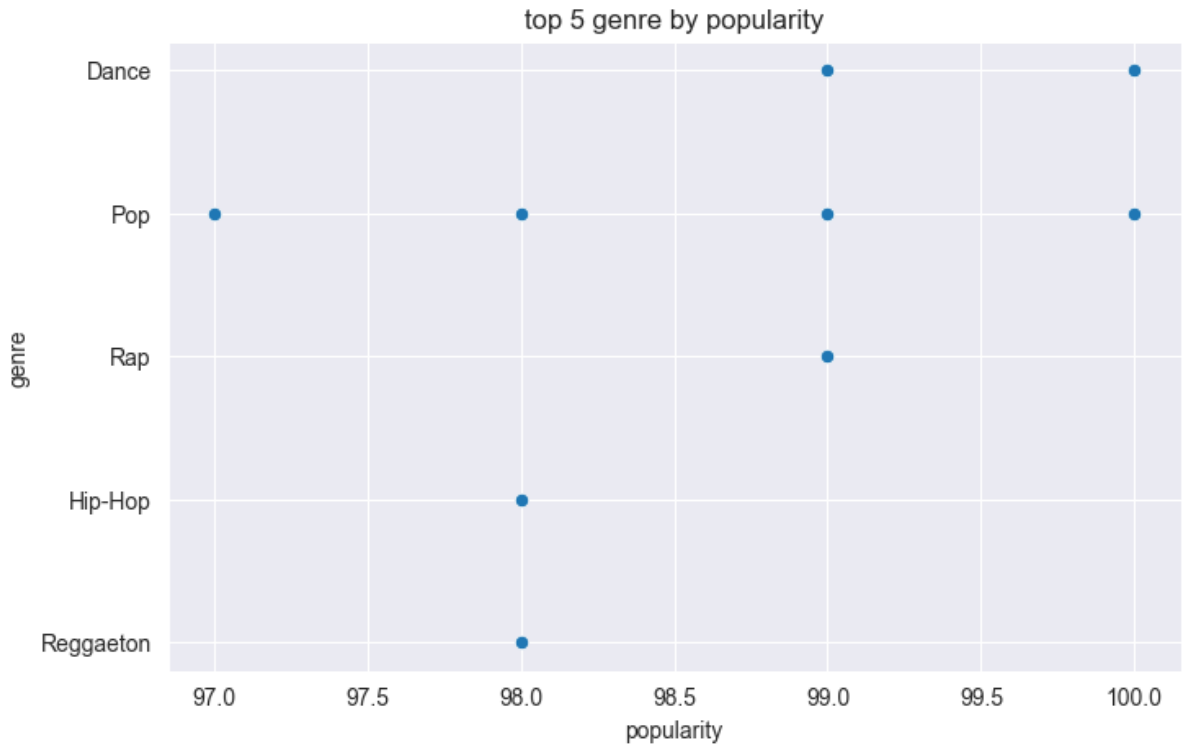
	genre	artist_name	track_name	track_id	popularity	acousticness	danceability	durati
0	Movie	Henri Salvador	C'est beau de faire un Show	0BRjO6ga9RKCKjfDqeFgWV	0	0.611	0.389	
1	Movie	Martin & les fées	Perdu d'avance (par Gad Elmaleh)	0BjC1NfoEOOusryehmNudP	1	0.246	0.590	1
2	Movie	Joseph Williams	Don't Let Me Be Lonely Tonight	0CoSDzoNIKCRs124s9uTVy	3	0.952	0.663	1
3	Movie	Henri Salvador	Dis-moi Monsieur Gordon Cooper	0Gc6TVm52BwZD07Ki6tIvf	0	0.703	0.240	1
4	Movie	Fabien Nataf	Ouverture	0IuslXpMROHdEPvSl1fTQK	4	0.950	0.331	




```
sns.set_style(style="darkgrid")
plt.figure(figsize=(8,5))
famous=a2.sort_values("popularity",ascending=False).head(10)
sns.barplot(y='artist_name',x="popularity",data=famous).set(title="top 5
artist by popularity")
```




```
sns.set_style(style="darkgrid")
plt.figure(figsize=(8,5))
famous=a2.sort_values("popularity",ascending=False).head(10)
sns.scatterplot(y='genre',x="popularity",data=famous).set(title="top 5 genre
by popularity")
```


top 5 genre by popularity
Dance
Pop
Rap
Hip-Hop
Reggaeton
genre
97.0
97.5
98.0
98.5
99.0
99.5
100.0
popularity