Tiktok

October 2, 2022

[1]:
```
import pandas as pd
import seaborn as sns
import matplotlib as plt
import numpy as py
```

[2]:
```
df = pd.read_csv(r'C:\Users\Damian\Desktop\Tiktok_songs.csv')
```

[3]:
```
df
```

[3]:

	track_name	artist_name	artist_pop	album	track_pop	danceability	energy	loudness	mode	key	speechiness	acousticness	instrumentalness	liveness	valence
0	Say So	Doja Cat	88	Hot Pink	80	0.787	0.673	-4.583	0	11	0.1590	0.26400	0.000003	0.0904	0.779
1	Blinding Lights	The Weeknd	93	After Hours	90	0.514	0.730	-5.934	1	1	0.0598	0.00146	0.000095	0.0897	0.334
2	Supalonely (feat. Gus Dapperton)	BENEE	67	Hey u x	63	0.862	0.631	-4.746	1						
3	Savage	Megan Thee Stallion	82	Suga	70	0.843	0.741	-5.609	1						
4	Moral of the Story	Ashe	68	Moral of the Story	76	0.572	0.406	-8.624	1						
..	…	…	…	…	…	…	…	…	…						
287	Buttons	The Pussycat Dolls	68	PCD	65	0.570	0.821	-4.380	1						
288	Get Busy	Sean Paul	79	Dutty Rock	74	0.735	0.824	-4.143	0						
289	ROCKSTAR (feat. Roddy Ricch)	DaBaby	82	BLAME IT ON BABY	80	0.746	0.690	-7.956	1						
290	Who Says	Selena Gomez & The Scene	67	When The Sun Goes Down	76	0.682	0.927	-2.915	1						
291	Crystal Dolphin	Engelwood	50	Crust FM	60	0.558	0.776	-6.868	1						

```
2      7       0.0515       0.29100          0.000209    0.1230    0.841
3     11       0.3340       0.02520          0.000000    0.0960    0.680
4     10       0.0427       0.58700          0.000004    0.1020    0.265
..   …          …           …                …           …         …
287    2       0.2670       0.17800          0.000000    0.2890    0.408
288   10       0.0360       0.61500          0.000000    0.1580    0.726
289   11       0.1640       0.24700          0.000000    0.1010    0.497
290    4       0.0479       0.08430          0.000000    0.1490    0.744
291    9       0.1790       0.33000          0.000445    0.4100    0.247

       tempo  time_signature  duration_ms
0    110.962               4       237893
1    171.005               4       200040
2    128.978               4       223488
3    168.983               4       155497
4    119.812               4       201084
..       …              …           …
287  210.857               4       225560
288  100.202               4       211666
289   89.977               4       181733
290  101.019               4       195613
291  128.064               4       114660

[292 rows x 18 columns]
```

[4]:
```
df_less = df.
 ↪drop(['key','mode','speechiness','acousticness','liveness','instrumentalness','time_signatu
 ↪axis=1)
```

[5]:
```
df_less.isnull().sum()
```

[5]:
```
track_name      0
artist_name     0
artist_pop      0
album           0
track_pop       0
danceability    0
energy          0
loudness        0
valence         0
tempo           0
duration_ms     0
dtype: int64
```

[6]:
```
df_less.describe()
```

[6]:
```
       artist_pop  track_pop  danceability      energy    loudness  \
count  292.000000  292.00000    292.000000  292.000000  292.000000
mean    67.318493   61.19863      0.725404    0.612855   -6.884469
std     16.571140   20.78498      0.143129    0.163366    2.697198
min      0.000000    0.00000      0.174000    0.045800  -23.928000
25%     57.750000   57.00000      0.651000    0.489250   -8.369750
50%     69.000000   66.00000      0.734500    0.631000   -6.523000
75%     80.000000   74.00000      0.831750    0.729000   -4.983500
max     95.000000   90.00000      0.980000    0.955000   -2.607000

          valence       tempo    duration_ms
count  292.000000  292.000000     292.000000
mean     0.545168  119.770538  192805.212329
std      0.212719   25.609372   55763.157757
min      0.038900   71.994000   37632.000000
25%      0.389750  100.064000  158673.000000
50%      0.554500  119.967000  191130.000000
75%      0.712750  131.334750  221394.750000
max      0.967000  210.857000  467587.000000
```

[7]:
```
df_less[df_less['duration_ms'] == max(df_less['duration_ms'])] #Najdłuższy utwór
```

[7]:
```
        track_name  artist_name  artist_pop     album  track_pop  \
256  Let It Happen  Tame Impala          78  Currents         74

     danceability  energy  loudness  valence    tempo  duration_ms
256         0.602   0.881    -5.875    0.577  125.012       467587
```

[8]:
```
df_less[df_less['duration_ms'] == min(df_less['duration_ms'])] #Najkrótszy utwór
```

[8]:
```
            track_name       artist_name  artist_pop                album  \
197  Knock at the Door  Playsongs People          25  Playsongs Rhyme Time

     track_pop  danceability  energy  loudness  valence    tempo  duration_ms
197         12         0.604   0.151   -23.928    0.216  209.082        37632
```

[9]:
```
df[['artist_name','track_name','duration_ms']].head(15)
```

[9]:
```
            artist_name                          track_name  duration_ms
0              Doja Cat                              Say So       237893
1            The Weeknd                      Blinding Lights       200040
2                 BENEE  Supalonely (feat. Gus Dapperton)       223488
3   Megan Thee Stallion                             Savage       155497
4                  Ashe                 Moral of the Story       201084
5             Lady Gaga   Rain On Me (with Ariana Grande)       182200
6             Lady Gaga                            Babylon       161733
7             Lady Gaga                             Replay       186653
8        Camila Cabello           My Oh My (feat. DaBaby)       170746
```

```
9              Dua Lipa                          Physical       193829
10                BENEE                           Glitter       180147
11          Ritt Momney                Put Your Records On       210463
12        Ariana Grande  Stuck with U (with Justin Bieber)       228482
13             Dua Lipa                      Break My Heart       221820
14         Madison Beer                            Selfish       223270
```

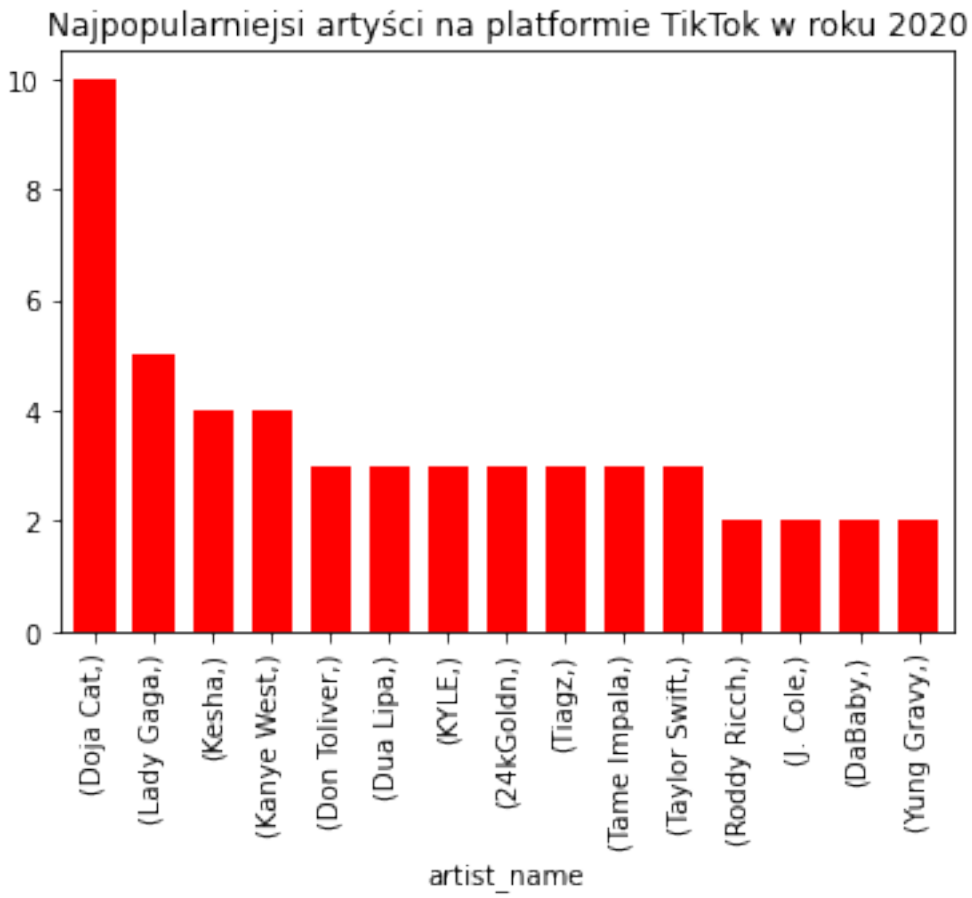
[10]:
```
df['duration_minutes'] = df['duration_ms']/1000/60
```

[11]:
```
Top15 = df[['artist_name','track_name','duration_minutes']].head(15)
Top15.describe()
```

[11]:
```
       duration_minutes
count         15.000000
mean           3.308161
std            0.425402
min            2.591617
25%            3.019558
50%            3.334000
75%            3.709083
max            3.964883
```

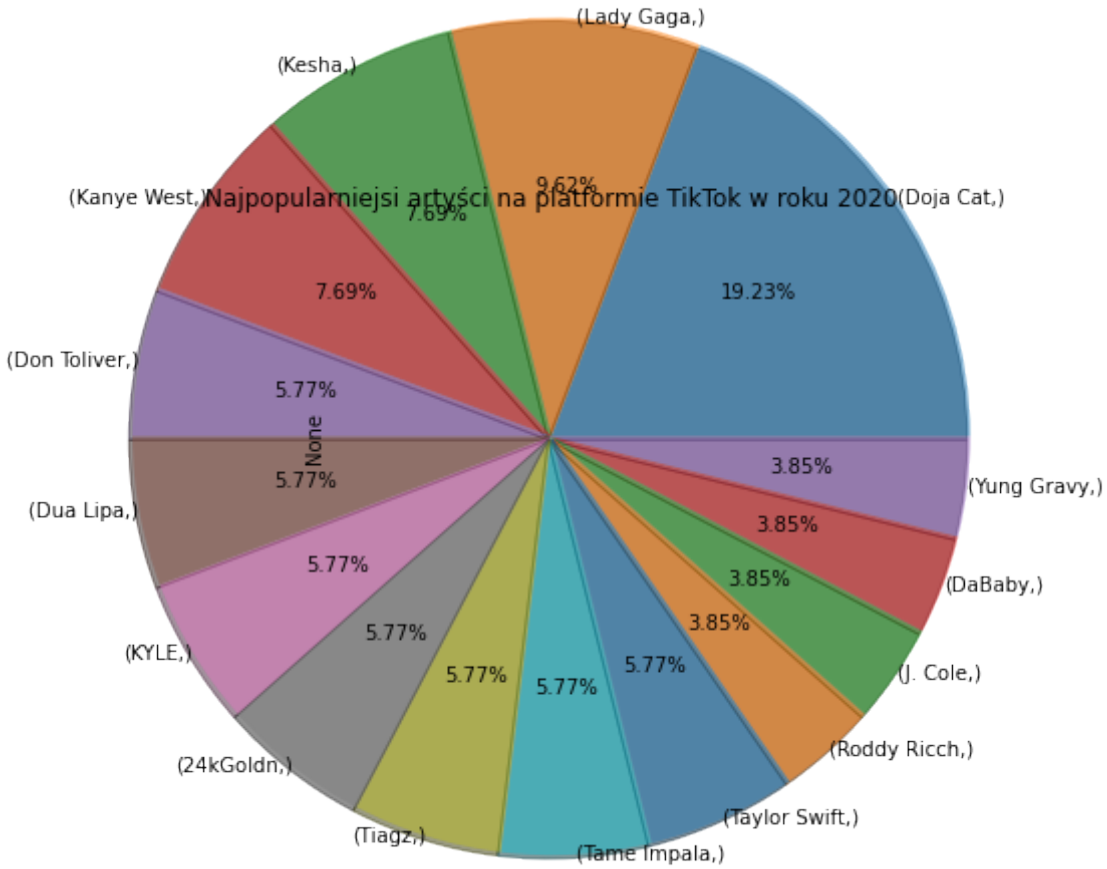
[12]:
```
df[['artist_name']].value_counts().head(15).plot(kind='bar',␣
↪title='Najpopularniejsi artyści na platformie TikTok w roku 2020', color =␣
↪'red', width = 0.7)
```

[12]:
```
<AxesSubplot:title={'center':'Najpopularniejsi artyści na platformie TikTok w
roku 2020'}, xlabel='artist_name'>
```



[13]:
```
df[['artist_name']].value_counts().head(15).plot(kind='pie',␣
↪title="Najpopularniejsi artyści na platformie TikTok w roku 2020",␣
↪shadow=True, radius = 2.4, labeldistance = 1, autopct = "%0.2f%%",␣
↪wedgeprops={'alpha':0.5})
```

[13]:
```
<AxesSubplot:title={'center':'Najpopularniejsi artyści na platformie TikTok w
roku 2020'}, ylabel='None'>
```


Najpopularniejsi artyści na platformie TikTok w roku 2020
(Lady Gaga,)
(Kesha,)
(Kanye West,)
(Doja Cat,)
9.62%
7.69%
7.69%
19.23%
(Don Toliver,)
5.77%
None
5.77%
3.85%
(Yung Gravy,)
(Dua Lipa,)
3.85%
(DaBaby,)
5.77%
3.85%
(KYLE,)
3.85%
5.77%
(J. Cole,)
5.77%
5.77%
5.77%
(Roddy Ricch,)
(24kGoldn,)
(Taylor Swift,)
(Tiagz,)
(Tame Impala,)