Spotify Musical Analysis

July 17, 2023

[1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

[2]:
```
music=pd.read_csv("datasets/spotify data.csv")
```

[3]:
```
music.head()
```

[3]:

	Column1	track_id	artists	album_name	track_name	popularity	Popularity Criteria's	duration_ms	explicit	danceability	…	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	time_signature	track_genre
0	0	5SuOikwiRyPMVoIQDJUgSV	Gen Hoshino	Comedy	Comedy	73	Most popular songs	230666	False	0.676	…	-6.746	0	0.1430	0.0322						
1	1	4qPNDBW1i3p13qLCt0Ki3A	Ben Woodward	Ghost (Acoustic)	Ghost - Acoustic	55	Popular song	149610	False	0.420	…	-17.235	1	0.0763	0.9240						
2	2	1iJBSr7s7jYXzM8EGcbK5b	Ingrid Michaelson;ZAYN	To Begin Again	To Begin Again	57	Popular song	210826	False	0.438	…	-9.734	1	0.0557	0.2100						
3	3	6lfxq3CG4xtTiEg7opyCyx	Kina Grannis	Crazy Rich Asians (Original Motion Picture Sou…	Can't Help Falling In Love	71	Most popular songs	201933	False	0.266	…	-18.515	1	0.0363	0.9050						
4	4	5vjLSffimiIP26QG5WcN2K	Chord Overstreet	Hold On	Hold On	82	Most popular songs	198853	False	0.618	…	-9.681	1	0.0526	0.4690						

```
0          0.000001    0.3580    0.715   87.917               4     acoustic
1          0.000006    0.1010    0.267   77.489               4     acoustic
2          0.000000    0.1170    0.120   76.332               4     acoustic
3          0.000071    0.1320    0.143  181.740               3     acoustic
4          0.000000    0.0829    0.167  119.949               4     acoustic

[5 rows x 22 columns]
```

[4]:
```
music.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 114000 entries, 0 to 113999
Data columns (total 22 columns):
 #   Column                 Non-Null Count   Dtype
---  ------                 --------------   -----
 0   Column1                114000 non-null  int64
 1   track_id               114000 non-null  object
 2   artists                113999 non-null  object
 3   album_name             113999 non-null  object
 4   track_name             113999 non-null  object
 5   popularity             114000 non-null  int64
 6   Popularity Criteria's  114000 non-null  object
 7   duration_ms            114000 non-null  int64
 8   explicit               114000 non-null  bool
 9   danceability           114000 non-null  float64
 10  energy                 114000 non-null  float64
 11  key                    114000 non-null  int64
 12  loudness               114000 non-null  float64
 13  mode                   114000 non-null  int64
 14  speechiness            114000 non-null  float64
 15  acousticness           114000 non-null  float64
 16  instrumentalness       114000 non-null  float64
 17  liveness               114000 non-null  float64
 18  valence                114000 non-null  float64
 19  tempo                  114000 non-null  float64
 20  time_signature         114000 non-null  int64
 21  track_genre            114000 non-null  object
dtypes: bool(1), float64(9), int64(6), object(6)
memory usage: 18.4+ MB
```

[7]:
```
music.describe()
```

[7]:
```
             Column1     popularity   duration_ms   danceability  \
count  114000.000000  114000.000000  1.140000e+05  114000.000000
mean    56999.500000      33.238535  2.280292e+05       0.566800
std     32909.109681      22.305078  1.072977e+05       0.173542
min         0.000000       0.000000  0.000000e+00       0.000000
25%     28499.750000      17.000000  1.740660e+05       0.456000
```

```
50%     56999.500000      35.000000  2.129060e+05       0.580000
75%     85499.250000      50.000000  2.615060e+05       0.695000
max    113999.000000     100.000000  5.237295e+06       0.985000

              energy            key       loudness           mode  \
count  114000.000000  114000.000000  114000.000000  114000.000000
mean        0.641383       5.309140      -8.258960       0.637553
std         0.251529       3.559987       5.029337       0.480709
min         0.000000       0.000000     -49.531000       0.000000
25%         0.472000       2.000000     -10.013000       0.000000
50%         0.685000       5.000000      -7.004000       1.000000
75%         0.854000       8.000000      -5.003000       1.000000
max         1.000000      11.000000       4.532000       1.000000

         speechiness   acousticness  instrumentalness       liveness  \
count  114000.000000  114000.000000     114000.000000  114000.000000
mean        0.084652       0.314910          0.156050       0.213553
std         0.105732       0.332523          0.309555       0.190378
min         0.000000       0.000000          0.000000       0.000000
25%         0.035900       0.016900          0.000000       0.098000
50%         0.048900       0.169000          0.000042       0.132000
75%         0.084500       0.598000          0.049000       0.273000
max         0.965000       0.996000          1.000000       1.000000

             valence          tempo  time_signature
count  114000.000000  114000.000000   114000.000000
mean        0.474068     122.147837        3.904035
std         0.259261      29.978197        0.432621
min         0.000000       0.000000        0.000000
25%         0.260000      99.218750        4.000000
50%         0.464000     122.017000        4.000000
75%         0.683000     140.071000        4.000000
max         0.995000     243.372000        5.000000
```

[5]:
```
music.isna().sum()
```

[5]:
```
Column1                 0
track_id                0
artists                 1
album_name              1
track_name              1
popularity              0
Popularity Criteria's   0
duration_ms             0
explicit                0
danceability            0
energy                  0
```

```
key                    0
loudness               0
mode                   0
speechiness            0
acousticness           0
instrumentalness       0
liveness               0
valence                0
tempo                  0
time_signature         0
track_genre            0
dtype: int64
```

[6]:
```
music=music.fillna("null")
```

[7]:
```
music.isna().sum()
```

[7]:
```
Column1                0
track_id               0
artists                0
album_name             0
track_name             0
popularity             0
Popularity Criteria's  0
duration_ms            0
explicit               0
danceability           0
energy                 0
key                    0
loudness               0
mode                   0
speechiness            0
acousticness           0
instrumentalness       0
liveness               0
valence                0
tempo                  0
time_signature         0
track_genre            0
dtype: int64
```

[8]:
```
total_track=music.track_id.nunique()
```

[9]:
```
print("Total No. Of Tracks are ",total_track)
```

```
Total No. Of Tracks are  89741
```

[10]:
```
total_artist=music.artists.nunique()
```

[11]:
```
print("Total No. Of Artist are ",total_artist)
```

```
Total No. Of Artist are  31438
```

[12]:
```
total_album=music.album_name.nunique()
```

[13]:
```
print("Total No. Of Album  are ",total_album)
```

```
Total No. Of Album  are  46580
```

[14]:
```
total_genre=music.track_genre.nunique()
```

[15]:
```
print("Total No. of Genre's are",total_genre)
```

```
Total No. of Genre's are 114
```

1 songs vs popularity criteria

[36]:
```
plt.style.available
```

[36]:
```
['Solarize_Light2',
 '_classic_test_patch',
 '_mpl-gallery',
 '_mpl-gallery-nogrid',
 'bmh',
 'classic',
 'dark_background',
 'fast',
 'fivethirtyeight',
 'ggplot',
 'grayscale',
 'seaborn-v0_8',
 'seaborn-v0_8-bright',
 'seaborn-v0_8-colorblind',
 'seaborn-v0_8-dark',
 'seaborn-v0_8-dark-palette',
 'seaborn-v0_8-darkgrid',
 'seaborn-v0_8-deep',
 'seaborn-v0_8-muted',
 'seaborn-v0_8-notebook',
 'seaborn-v0_8-paper',
 'seaborn-v0_8-pastel',
 'seaborn-v0_8-poster',
 'seaborn-v0_8-talk',
 'seaborn-v0_8-ticks',
 'seaborn-v0_8-white',
 'seaborn-v0_8-whitegrid',
 'tableau-colorblind10']
```

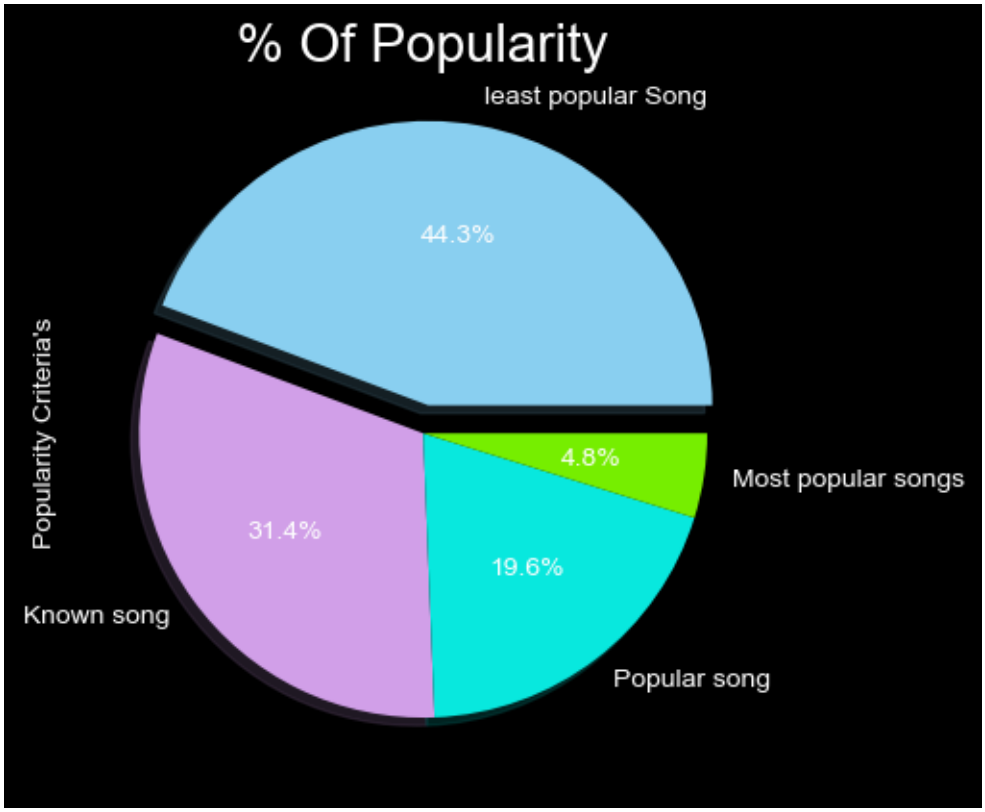
[71]:
```
plt.style.use('seaborn-v0_8-bright')
```

[54]:
```
```

[55]:
```
music["Popularity Criteria's"].value_counts()
```

[55]:
```
least popular Song    50487
Known song            35743
Popular song          22298
Most popular songs     5472
Name: Popularity Criteria's, dtype: int64
```

[72]:
```
music["Popularity Criteria's"].value_counts().sort_values(ascending=False).
 ↪plot(kind="pie",
 ↪autopct="%1.1f%%",
 ↪colors=["#89CFF0","#D19FE8","#08E8DE","#76EE00"],
 ↪explode=(0.1,0,0,0),
 ↪shadow=True)
plt.title("% Of Popularity",size=20)
plt.show()
```



2 top 5 most no of songs by artists

[18]:
```
music.columns
```

[18]:
```
Index(['Column1', 'track_id', 'artists', 'album_name', 'track_name',
       'popularity', 'Popularity Criteria's', 'duration_ms', 'explicit',
       'danceability', 'energy', 'key', 'loudness', 'mode', 'speechiness',
       'acousticness', 'instrumentalness', 'liveness', 'valence', 'tempo',
       'time_signature', 'track_genre'],
      dtype='object')
```

[19]:
```
music.artists.value_counts().head(5)
```

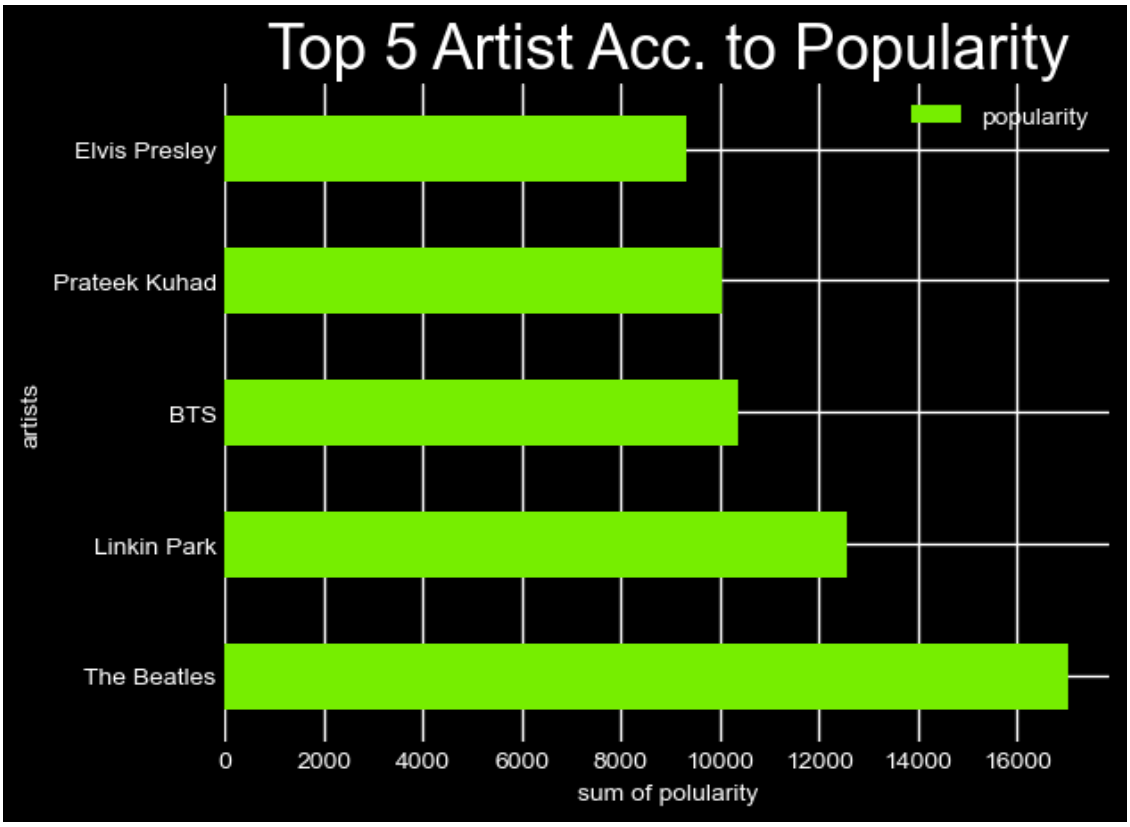
[19]:
```
The Beatles        279
George Jones       271
Stevie Wonder      236
Linkin Park        224
Ella Fitzgerald    222
Name: artists, dtype: int64
```

3 top 5 artist most popular by popularity

[20]:
```
music[["artists","popularity"]].groupby("artists").sum("popularity").
 ↪sort_values("popularity",ascending=False).head(5)
```

[20]:
```
               popularity
artists
The Beatles         17021
Linkin Park         12560
BTS                 10358
Prateek Kuhad       10054
Elvis Presley        9307
```

[88]:
```
music[["artists","popularity"]].groupby("artists").sum("popularity").
 ↪sort_values("popularity",ascending=False).head(5).
 ↪plot(kind="barh",color="#76EE00")
plt.xlabel("sum of polularity")
plt.title("Top 5 Artist Acc. to Popularity", size=25)
plt.show()
```



4 Top 5 loudest tracks

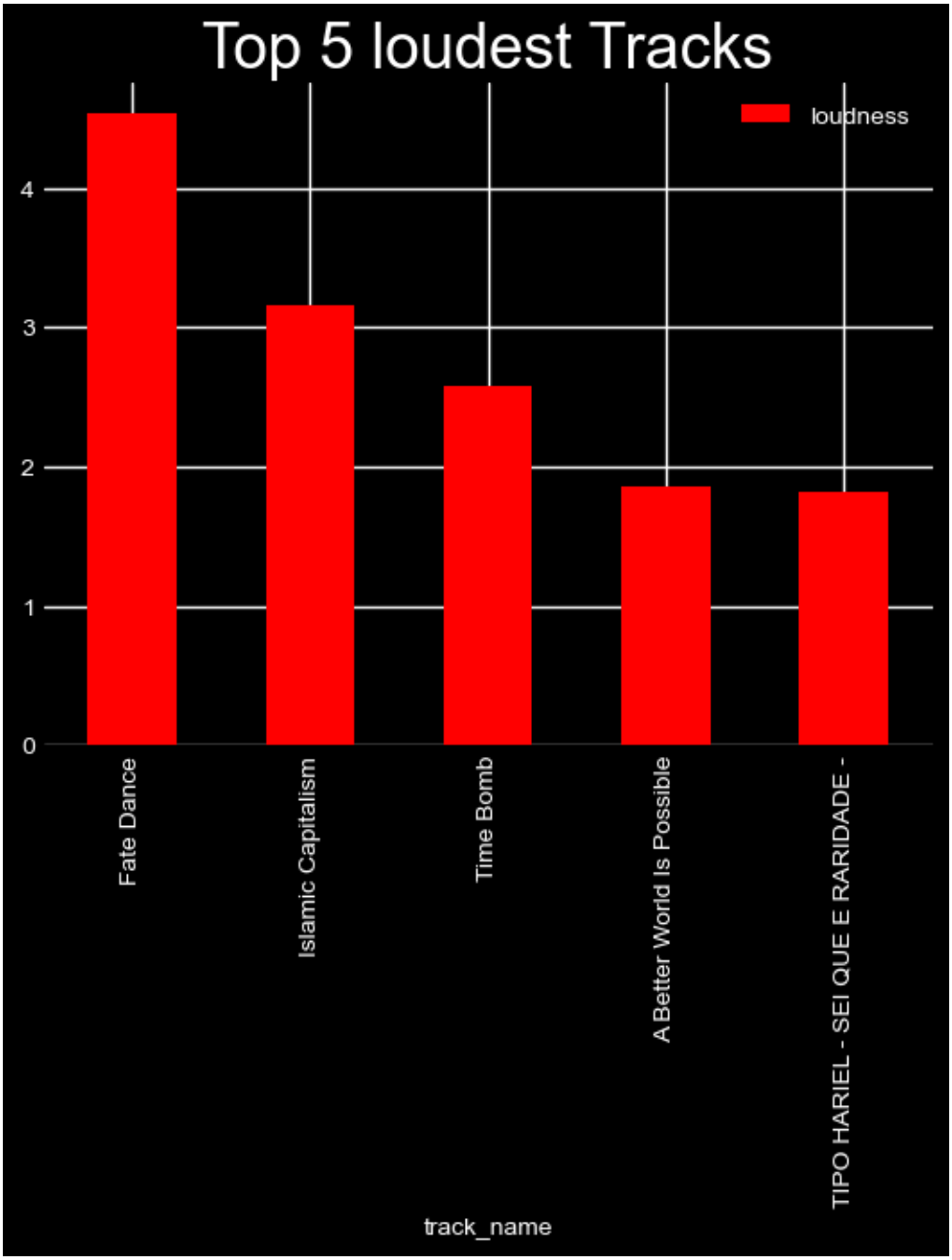
[61]:
```
music[["loudness","track_name"]].sort_values(by="loudness",ascending=False).
↪head(5).set_index("track_name")
```

[61]:

track_name	loudness
Fate Dance	4.532
Islamic Capitalism	3.156
Time Bomb	2.574
A Better World Is Possible	1.864
TIPO HARIEL - SEI QUE E RARIDADE -	1.821

[90]:
```
music[["loudness","track_name"]].sort_values(by="loudness",ascending=False).
↪set_index("track_name").head(5).plot(kind="bar",color="#FF0000")
plt.title("Top 5 loudest Tracks",size=25)
```

[90]:
```
Text(0.5, 1.0, 'Top 5 loudest Tracks')
```



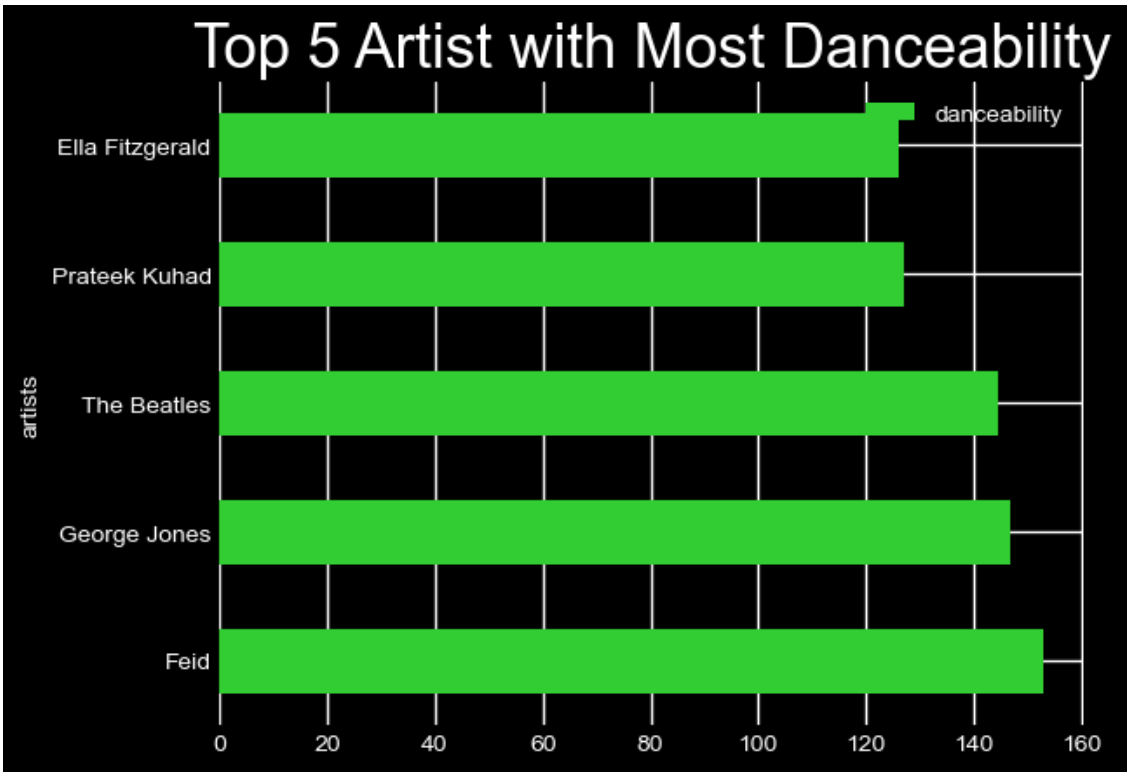
5 artist with most danceability song

[26]:
```
danceability=music[["artists","danceability"]].groupby("artists").
  ↪sum("danceability").sort_values(by="danceability",ascending=False).head(5)
danceability
```

[26]:
```
                 danceability
artists
Feid                  152.837
George Jones          146.655
The Beatles           144.649
Prateek Kuhad         126.868
Ella Fitzgerald       125.940
```

[91]:
```
danceability.plot(kind="barh",color="#32CD32")
plt.title("Top 5 Artist with Most Danceability",size=25)
```

[91]:
```
Text(0.5, 1.0, 'Top 5 Artist with Most Danceability')
```



6 top 10 instumental tracks

[63]:
```
music[['instrumentalness',"track_name"]].
  ↪sort_values(by="instrumentalness",ascending=False).head(10).
  ↪set_index("track_name")
```

[63]:
```
                                    instrumentalness
track_name
Brown Sleep Noise                                1.0
```

```
Ruido Rosa Puro - Una Hora Versión                    1.0
Brown Noise 148 LPF                                  1.0
Fan Whir                                             1.0
Pink Noise For Sleep 8                               1.0
Fan Whir                                             1.0
Clean White Noise - Loopable with no fade            1.0
Air Conditioner Sound                                1.0
Brown Noise For Sleep 5                              1.0
Hair Dryer                                           1.0
```

7 keys

[31]:
```
keys=music.key.nunique()
```

[32]:
```
print("Total No. Of Keys are ",keys)
```

```
Total No. Of Keys are  12
```

8 most listen key

[33]:
```
music.key.value_counts()
```

[33]:
```
7     13245
0     13061
2     11644
9     11313
1     10772
5      9368
11     9282
4      9008
6      7921
10     7456
8      7360
3      3570
Name: key, dtype: int64
```

9 genre top5 most popular

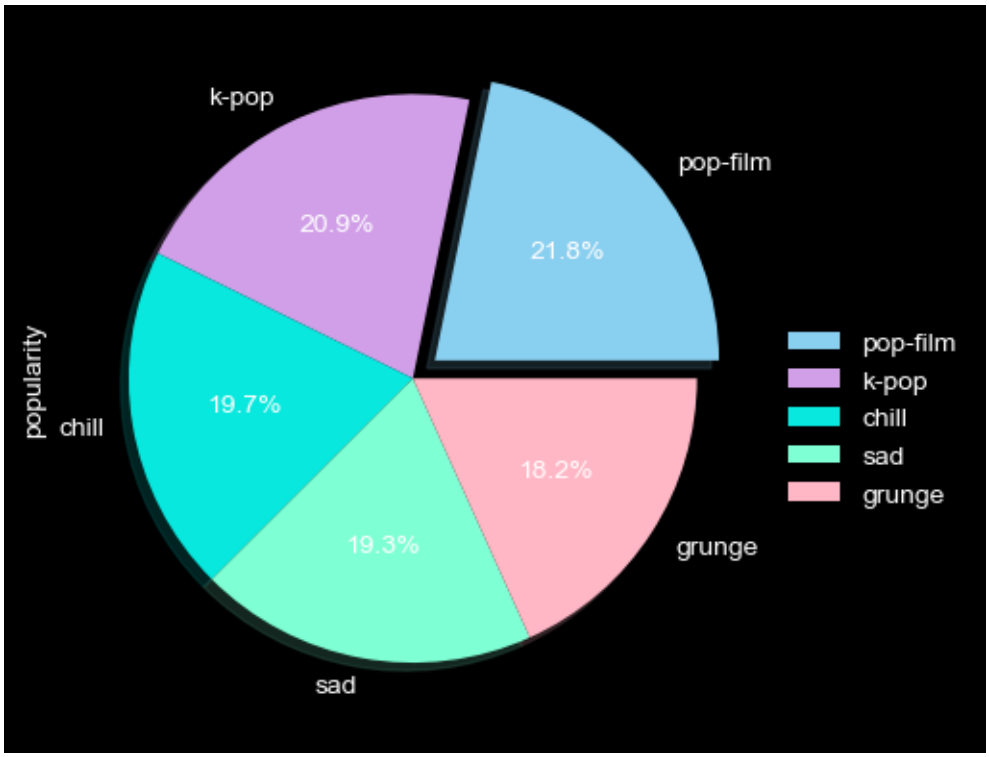
[34]:
```
music[["track_genre","popularity"]].groupby("track_genre").sum("popularity").
↪sort_values(by="popularity",ascending=False).head(5)
```

[34]:
```
             popularity
track_genre
pop-film          59283
k-pop             56896
```

```
chill           53651
sad             52379
grunge          49594
```

[85]:
```
music[["track_genre","popularity"]].groupby("track_genre").sum("popularity").
  ↪sort_values(by="popularity",ascending=False).head(5).
  ↪plot(kind="pie",y="popularity",autopct="%1.
  ↪1f%%",colors=["#89CFF0","#D19FE8","#08E8DE","#7FFFD4","#FFB7C5"],explode=(0.
  ↪1,0,0,0,0),shadow=True)
plt.legend(bbox_to_anchor=(1.3, 0.6))
```

[85]:
```
<matplotlib.legend.Legend at 0x25dfa9be080>
```



[ ]: