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

```
Matplotlib is building the font cache; this may take a moment.
```

```
Sdf = pd.read_csv(r"C:\New folder\spotify-2023.csv", encoding='ISO-8859-1')
```

```
Sdf.head()
```

```
                            track_name    artist(s)_name  artist_count
\
0  Seven (feat. Latto) (Explicit Ver.)  Latto, Jung Kook             2

1                                 LALA       Myke Towers             1

2                              vampire    Olivia Rodrigo             1

3                         Cruel Summer      Taylor Swift             1

4                       WHERE SHE GOES         Bad Bunny             1


   released_year  released_month  released_day
in_spotify_playlists  \
0           2023               7            14                    553

1           2023               3            23                   1474

2           2023               6            30                   1397

3           2019               8            23                   7858

4           2023               5            18                   3133


   in_spotify_charts    streams  in_apple_playlists  ...  bpm key
mode  \
0                147  141381703                  43  ...  125   B
Major
1                 48  133716286                  48  ...   92  C#
Major
2                113  140003974                  94  ...  138   F
Major
3                100  800840817                 116  ...  170   A
Major
4                 50  303236322                  84  ...  144   A
Minor
```

```
  danceability_%  valence_% energy_% acousticness_%  instrumentalness_
%  \
0              80         89       83              31
0
1              71         61       74               7
0
2              51         32       53              17
0
3              55         58       72              11
0
4              65         23       80              14
63

   liveness_%  speechiness_%
0           8              4
1          10              4
2          31              6
3          11             15
4          11              6

[5 rows x 24 columns]
```

```
Sdf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 953 entries, 0 to 952
Data columns (total 24 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   track_name            953 non-null    object
 1   artist(s)_name        953 non-null    object
 2   artist_count          953 non-null    int64
 3   released_year         953 non-null    int64
 4   released_month        953 non-null    int64
 5   released_day          953 non-null    int64
 6   in_spotify_playlists  953 non-null    int64
 7   in_spotify_charts     953 non-null    int64
 8   streams               953 non-null    object
 9   in_apple_playlists    953 non-null    int64
 10  in_apple_charts       953 non-null    int64
 11  in_deezer_playlists   953 non-null    object
 12  in_deezer_charts      953 non-null    int64
 13  in_shazam_charts      903 non-null    object
 14  bpm                   953 non-null    int64
 15  key                   858 non-null    object
 16  mode                  953 non-null    object
 17  danceability_%        953 non-null    int64
 18  valence_%             953 non-null    int64
 19  energy_%              953 non-null    int64
 20  acousticness_%        953 non-null    int64
```

```
 21  instrumentalness_%    953 non-null    int64
 22  liveness_%            953 non-null    int64
 23  speechiness_%         953 non-null    int64
dtypes: int64(17), object(7)
memory usage: 178.8+ KB
```

```
Sdf.isnull()
```

```
     track_name  artist(s)_name  artist_count  released_year
released_month  \
0         False           False         False          False
False
1         False           False         False          False
False
2         False           False         False          False
False
3         False           False         False          False
False
4         False           False         False          False
False
..          ...             ...           ...            ...
...
948       False           False         False          False
False
949       False           False         False          False
False
950       False           False         False          False
False
951       False           False         False          False
False
952       False           False         False          False
False

     released_day  in_spotify_playlists  in_spotify_charts  streams  \
0           False                 False              False    False
1           False                 False              False    False
2           False                 False              False    False
3           False                 False              False    False
4           False                 False              False    False
..            ...                   ...                ...      ...
948         False                 False              False    False
949         False                 False              False    False
950         False                 False              False    False
951         False                 False              False    False
952         False                 False              False    False

     in_apple_playlists  ...    bpm    key   mode  danceability_%
valence_%  \
0                 False  ...  False  False  False           False
False
```

```
1                False  ...  False  False  False           False
False
2                False  ...  False  False  False           False
False
3                False  ...  False  False  False           False
False
4                False  ...  False  False  False           False
False
..                 ...  ...    ...    ...    ...             ...
...
948              False  ...  False  False  False           False
False
949              False  ...  False  False  False           False
False
950              False  ...  False  False  False           False
False
951              False  ...  False  False  False           False
False
952              False  ...  False  False  False           False
False

     energy_%  acousticness_%  instrumentalness_%  liveness_%
speechiness_%
0       False           False               False       False
False
1       False           False               False       False
False
2       False           False               False       False
False
3       False           False               False       False
False
4       False           False               False       False
False
..        ...             ...                 ...         ...
...
948     False           False               False       False
False
949     False           False               False       False
False
950     False           False               False       False
False
951     False           False               False       False
False
952     False           False               False       False
False

[953 rows x 24 columns]

Sdf.isnull().sum()
```

```
track_name                0
artist(s)_name            0
artist_count              0
released_year             0
released_month            0
released_day              0
in_spotify_playlists      0
in_spotify_charts         0
streams                   0
in_apple_playlists        0
in_apple_charts           0
in_deezer_playlists       0
in_deezer_charts          0
in_shazam_charts         50
bpm                       0
key                      95
mode                      0
danceability_%            0
valence_%                 0
energy_%                  0
acousticness_%            0
instrumentalness_%        0
liveness_%                0
speechiness_%             0
dtype: int64
```

```
Sdf.dropna(inplace = True)
```

```
Sdf.isnull().sum()
```

```
track_name              0
artist(s)_name          0
artist_count            0
released_year           0
released_month          0
released_day            0
in_spotify_playlists    0
in_spotify_charts       0
streams                 0
in_apple_playlists      0
in_apple_charts         0
in_deezer_playlists     0
in_deezer_charts        0
in_shazam_charts        0
bpm                     0
key                     0
mode                    0
danceability_%          0
valence_%               0
energy_%                0
```

```
acousticness_%          0
instrumentalness_%      0
liveness_%              0
speechiness_%           0
dtype: int64
```

```
Sdf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 817 entries, 0 to 952
Data columns (total 24 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   track_name            817 non-null    object
 1   artist(s)_name        817 non-null    object
 2   artist_count          817 non-null    int64
 3   released_year         817 non-null    int64
 4   released_month        817 non-null    int64
 5   released_day          817 non-null    int64
 6   in_spotify_playlists  817 non-null    int64
 7   in_spotify_charts     817 non-null    int64
 8   streams               817 non-null    object
 9   in_apple_playlists    817 non-null    int64
 10  in_apple_charts       817 non-null    int64
 11  in_deezer_playlists   817 non-null    object
 12  in_deezer_charts      817 non-null    int64
 13  in_shazam_charts      817 non-null    object
 14  bpm                   817 non-null    int64
 15  key                   817 non-null    object
 16  mode                  817 non-null    object
 17  danceability_%        817 non-null    int64
 18  valence_%             817 non-null    int64
 19  energy_%              817 non-null    int64
 20  acousticness_%        817 non-null    int64
 21  instrumentalness_%    817 non-null    int64
 22  liveness_%            817 non-null    int64
 23  speechiness_%         817 non-null    int64
dtypes: int64(17), object(7)
memory usage: 159.6+ KB
```

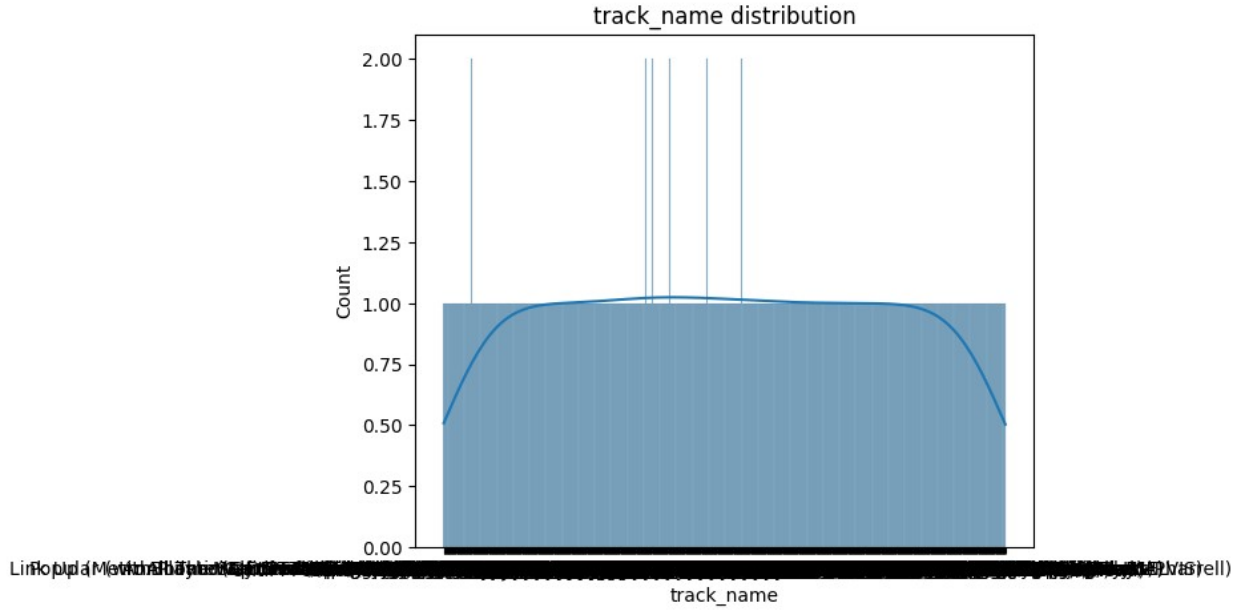
```
Sdf.describe(include = 'all')
```

```
         track_name artist(s)_name  artist_count  released_year  \
count           817            817    817.000000     817.000000
unique          811            571           NaN            NaN
top     Die For You   Taylor Swift           NaN            NaN
freq              2             29           NaN            NaN
mean            NaN            NaN      1.567931    2018.457772
std             NaN            NaN      0.876211      10.829267
min             NaN            NaN      1.000000    1930.000000
```

```
25%              NaN             NaN      1.000000    2021.000000
50%              NaN             NaN      1.000000    2022.000000
75%              NaN             NaN      2.000000    2022.000000
max              NaN             NaN      8.000000    2023.000000

        released_month  released_day  in_spotify_playlists
in_spotify_charts  \
count       817.000000    817.000000            817.000000
817.000000
unique             NaN           NaN                   NaN
NaN
top                NaN           NaN                   NaN
NaN
freq               NaN           NaN                   NaN
NaN
mean          6.018360     13.696450           4849.898409
11.722154
std           3.572554      9.299663           7741.126455
18.617668
min           1.000000      1.000000             31.000000
0.000000
25%           3.000000      5.000000            829.000000
0.000000
50%           5.000000     13.000000           2040.000000
3.000000
75%           9.000000     22.000000           4890.000000
16.000000
max          12.000000     31.000000          52898.000000
147.000000

           streams  in_apple_playlists  ...         bpm  key    mode  \
count          817          817.000000  ...  817.000000  817     817
unique         814                 NaN  ...         NaN   11       2
top     1223481149                 NaN  ...         NaN   C#  Major
freq             2                 NaN  ...         NaN  115     452
mean           NaN           60.161567  ...  122.565483  NaN     NaN
std            NaN           74.923594  ...   28.174803  NaN     NaN
min            NaN            0.000000  ...   65.000000  NaN     NaN
25%            NaN           12.000000  ...   99.000000  NaN     NaN
50%            NaN           32.000000  ...  120.000000  NaN     NaN
75%            NaN           78.000000  ...  141.000000  NaN     NaN
max            NaN          532.000000  ...  206.000000  NaN     NaN

       danceability_%   valence_%    energy_% acousticness_%  \
count      817.000000  817.000000  817.000000     817.000000
unique            NaN         NaN         NaN            NaN
top               NaN         NaN         NaN            NaN
freq              NaN         NaN         NaN            NaN
mean        67.391677   51.201958   64.362301      26.309670
std         14.688458   23.620978   16.107587      25.470972
```

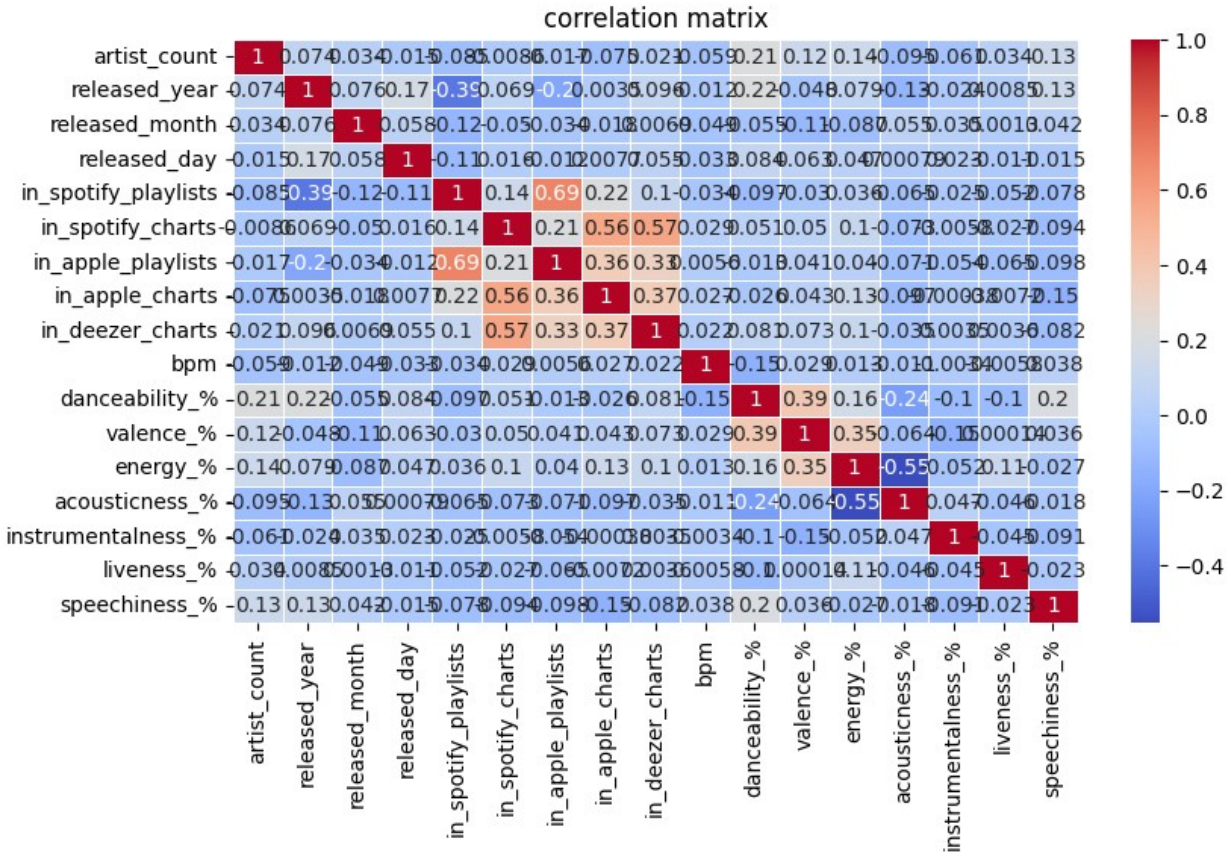
```
min          23.000000    4.000000   14.000000       0.000000
25%          57.000000   32.000000   53.000000       5.000000
50%          70.000000   51.000000   66.000000      17.000000
75%          79.000000   70.000000   76.000000      41.000000
max          96.000000   97.000000   97.000000      97.000000

        instrumentalness_%  liveness_%  speechiness_%
count           817.000000  817.000000     817.000000
unique                 NaN         NaN            NaN
top                    NaN         NaN            NaN
freq                   NaN         NaN            NaN
mean              1.676867   18.168911      10.526316
std               8.767328   13.541996      10.219987
min               0.000000    3.000000       2.000000
25%               0.000000   10.000000       4.000000
50%               0.000000   12.000000       6.000000
75%               0.000000   24.000000      12.000000
max              91.000000   97.000000      64.000000

[11 rows x 24 columns]
```

```
plt.figure(figsize=(6,5))
sns.histplot(Sdf['track_name'],kde=True,bins=10)
plt.title("track_name distribution")
plt.show()
```

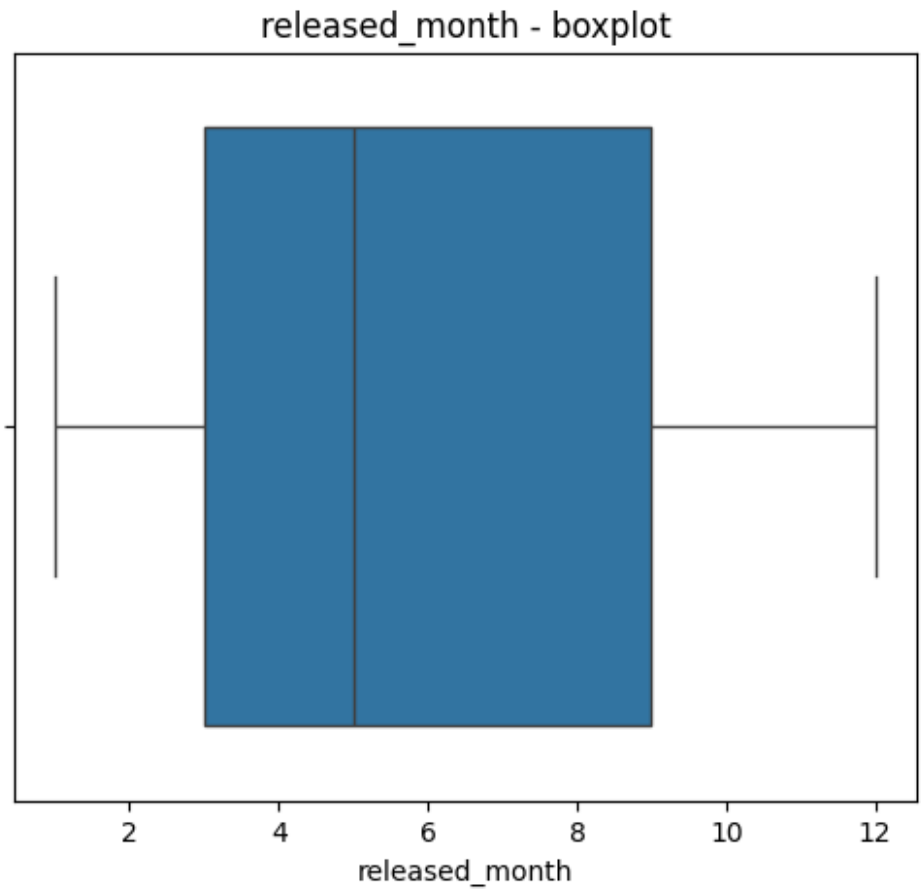



```
Sdf_num = Sdf.select_dtypes(include=['number'])
Sdf_num.head()
```

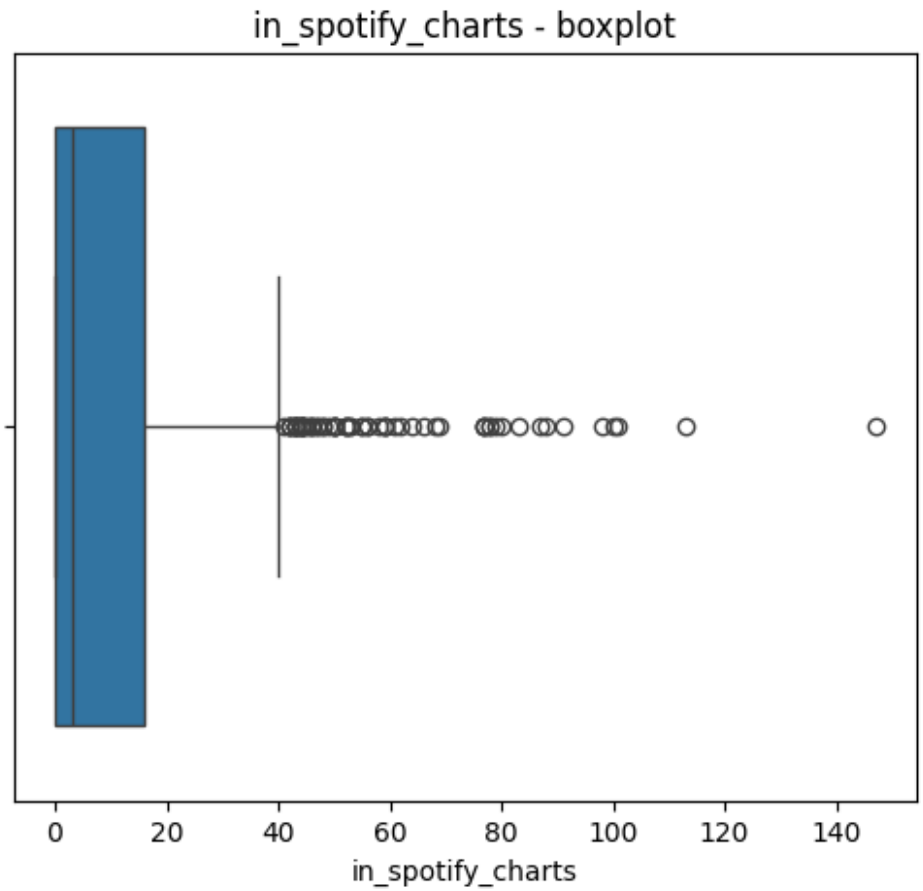
```
   artist_count  released_year  released_month  released_day  \
0             2           2023               7            14
1             1           2023               3            23
2             1           2023               6            30
3             1           2019               8            23
4             1           2023               5            18

   in_spotify_playlists  in_spotify_charts  in_apple_playlists  \
0                   553                147                  43
1                  1474                 48                  48
2                  1397                113                  94
3                  7858                100                 116
4                  3133                 50                  84

   in_apple_charts  in_deezer_charts  bpm  danceability_%  valence_
%  \
0              263                10  125              80         89

1              126                14   92              71         61

2              207                14  138              51         32

3              207                12  170              55         58

4              133                15  144              65         23


   energy_%  acousticness_%  instrumentalness_%  liveness_%
speechiness_%
0        83              31                   0           8
4
1        74               7                   0          10
4
2        53              17                   0          31
6
3        72              11                   0          11
15
4        80              14                  63          11
6

plt.figure(figsize=(9,5))
sns.heatmap(Sdf_num.corr(),annot=True,cmap='coolwarm',linewidth=0.5)
plt.title("correlation matrix")
plt.show()
```



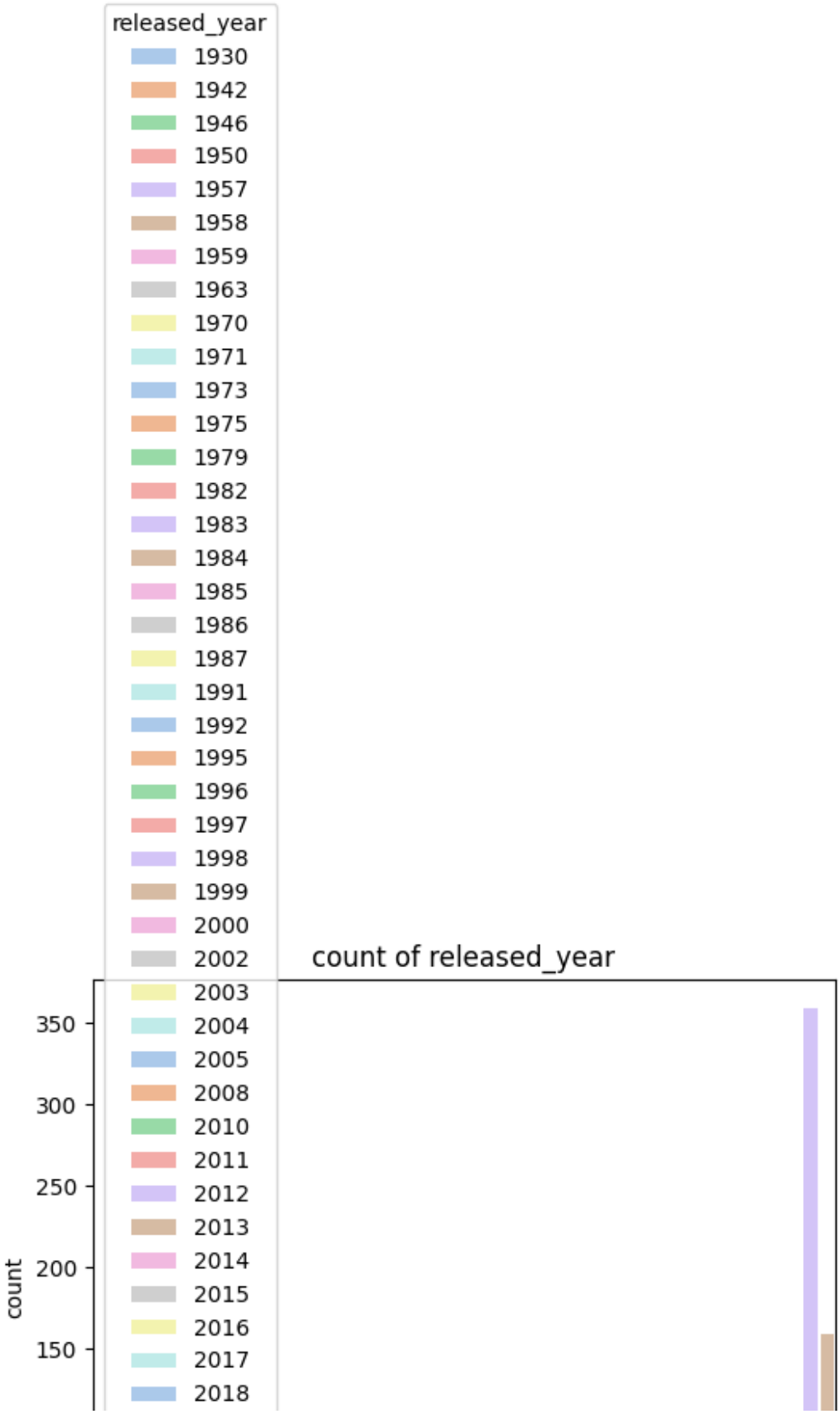

```
plt.figure(figsize=(6,5))
sns.boxplot(x=Sdf['released_month'])
plt.title("released_month - boxplot")
plt.show()
```



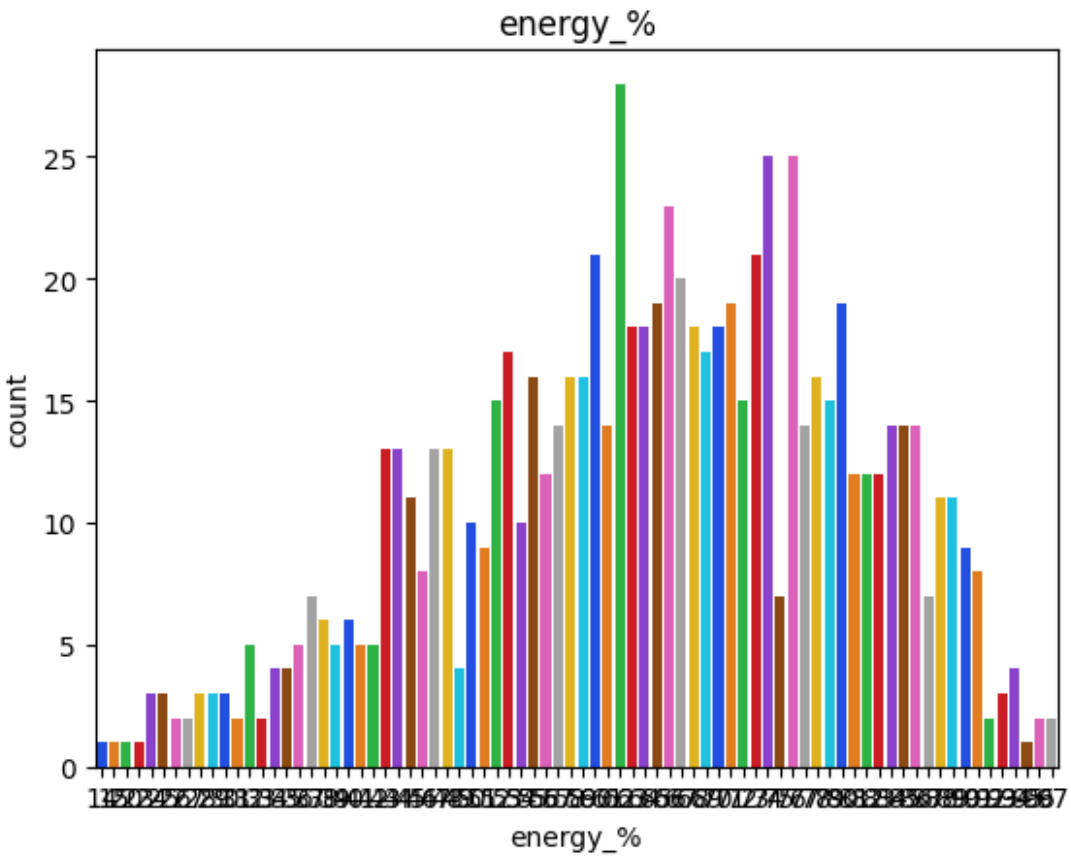

```
plt.figure(figsize=(6,5))
sns.boxplot(x=Sdf['in_spotify_charts'])
plt.title("in_spotify_charts - boxplot")
plt.show()
```




```
plt.figure(figsize=(6,5))
sns.countplot(x=Sdf['released_year'],hue =
Sdf['released_year'],palette='pastel')
plt.title('count of released_year')
plt.show()
```

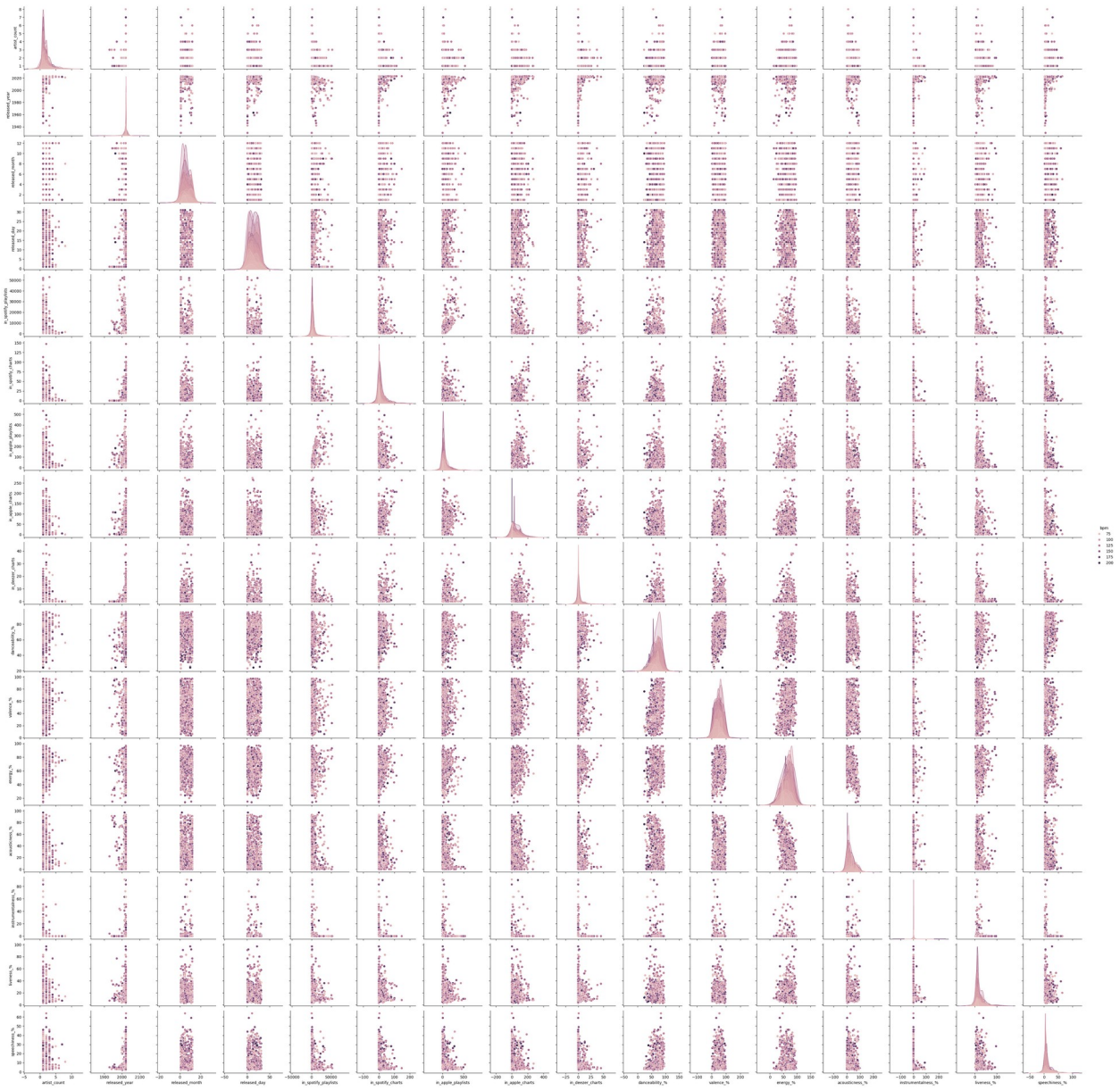

released_year
1930
1942
1946
1950
1957
1958
1959
1963
1970
1971
1973
1975
1979
1982
1983
1984
1985
1986
1987
1991
1992
1995
1996
1997
1998
1999
2000
2002
2003
2004
2005
2008
2010
2011
2012
2013
2014
2015
2016
2017
2018
count of released_year
count
350
300
250
200
150

```
sns.countplot(x=Sdf['energy_%'],hue=Sdf['energy_%'],palette =
'bright',legend = False)
plt.title("energy_%")
plt.show()
```




```
sns.pairplot(Sdf,hue='bpm')
```

```
<seaborn.axisgrid.PairGrid at 0x155ceda63f0>
```



```
plt.show()

Sdf.groupby('bpm')['danceability_%'].mean()

bpm
65     71.0
67     53.0
71     59.0
72     60.0
73     83.0
       ...
198    59.0
200    39.0
202    29.0
```

```
204    52.0
206    43.0
Name: danceability_%, Length: 122, dtype: float64
```