```
import numpy as np
import pandas as pd

df=pd.read_csv(r'/home/rohit/csv files/spotify-
2023.csv',encoding='ISO-8859-1')

df
```

```
                                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
..                                   ...                 ...
...
948                          My Mind & Me        Selena Gomez
1
949             Bigger Than The Whole Sky        Taylor Swift
1
950                  A Veces (feat. Feid)  Feid, Paulo Londra
2
951                          En La De Ella  Feid, Sech, Jhayco
3
952                                 Alone           Burna Boy
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

..             ...             ...           ...                   ...

948           2022              11             3                   953

949           2022              10            21                  1180
```

```
950            2022              11              3                    573

951            2022              10             20                   1320

952            2022              11              4                    782


     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
..                 ...        ...                 ...  ...  ...  ..
...
948                  0   91473363                  61  ...  144   A
Major
949                  0  121871870                   4  ...  166  F#
Major
950                  0   73513683                   2  ...   92  C#
Major
951                  0  133895612                  29  ...   97  C#
Major
952                  2   96007391                  27  ...   90   E
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
..              ...        ...      ...             ...
...
948              60         24       39              57
0
949              42          7       24              83
1
```

```
950              80          81        67                4
0
951              82          67        77                8
0
952              61          32        67               15
0

     liveness_%  speechiness_%
0             8              4
1            10              4
2            31              6
3            11             15
4            11              6
..          ...            ...
948           8              3
949          12              6
950           8              6
951          12              5
952          11              5

[953 rows x 24 columns]
```

```
df.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
```

```
x=['track_name','artist(s)_name','artist_count','released_year','in_spotify_playlists','in_spotify_charts','streams']

data=df[x]

data
```

```
                                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
..                                   ...                 ...
...
948                        My Mind & Me        Selena Gomez
1
949            Bigger Than The Whole Sky        Taylor Swift
1
950                A Veces (feat. Feid)  Feid, Paulo Londra
2
951                       En La De Ella  Feid, Sech, Jhayco
3
952                              Alone           Burna Boy
1

     released_year  in_spotify_playlists  in_spotify_charts    streams
0             2023                   553                147  141381703
1             2023                  1474                 48  133716286
2             2023                  1397                113  140003974
3             2019                  7858                100  800840817
4             2023                  3133                 50  303236322
..             ...                   ...                ...        ...
948           2022                   953                  0   91473363
949           2022                  1180                  0  121871870
950           2022                   573                  0   73513683
```

```
951           2022                   1320                    0  133895612

952           2022                    782                    2   96007391


[953 rows x 7 columns]
```

```
import matplotlib.pyplot as plt
import seaborn as sns

data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 953 entries, 0 to 952
Data columns (total 7 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   in_spotify_playlists  953 non-null    int64 
 5   in_spotify_charts     953 non-null    int64 
 6   streams               953 non-null    object
dtypes: int64(4), object(3)
memory usage: 52.2+ KB
```

```
df.describe()
```

```
       artist_count  released_year  released_month  released_day  \
count    953.000000     953.000000      953.000000    953.000000   
mean       1.556139    2018.238195        6.033578     13.930745   
std        0.893044      11.116218        3.566435      9.201949   
min        1.000000    1930.000000        1.000000      1.000000   
25%        1.000000    2020.000000        3.000000      6.000000   
50%        1.000000    2022.000000        6.000000     13.000000   
75%        2.000000    2022.000000        9.000000     22.000000   
max        8.000000    2023.000000       12.000000     31.000000   

       in_spotify_playlists  in_spotify_charts  in_apple_playlists  \
count            953.000000         953.000000          953.000000   
mean            5200.124869          12.009444           67.812172   
std             7897.608990          19.575992           86.441493   
min               31.000000           0.000000            0.000000   
25%              875.000000           0.000000           13.000000   
50%             2224.000000           3.000000           34.000000   
75%             5542.000000          16.000000           88.000000   
max            52898.000000         147.000000          672.000000   
```

```
       in_apple_charts  in_deezer_charts         bpm  danceability_
%  \
count       953.000000        953.000000  953.000000       953.00000
mean         51.908709          2.666317  122.540399        66.96957
std          50.630241          6.035599   28.057802        14.63061
min           0.000000          0.000000   65.000000        23.00000
25%           7.000000          0.000000  100.000000        57.00000
50%          38.000000          0.000000  121.000000        69.00000
75%          87.000000          2.000000  140.000000        78.00000
max         275.000000         58.000000  206.000000        96.00000


        valence_%    energy_%  acousticness_%  instrumentalness_%
liveness_%  \
count  953.000000  953.000000      953.000000          953.000000
953.000000
mean    51.431270   64.279119       27.057712            1.581322
18.213012
std     23.480632   16.550526       25.996077            8.409800
13.711223
min      4.000000    9.000000        0.000000            0.000000
3.000000
25%     32.000000   53.000000        6.000000            0.000000
10.000000
50%     51.000000   66.000000       18.000000            0.000000
12.000000
75%     70.000000   77.000000       43.000000            0.000000
24.000000
max     97.000000   97.000000       97.000000           91.000000
97.000000

       speechiness_%
count     953.000000
mean       10.131165
std         9.912888
min         2.000000
25%         4.000000
50%         6.000000
75%        11.000000
max        64.000000
```

```
# Remove rows with null values in the 'streams' column
data = data.dropna(subset=['streams'])
```

```python
# Reset the index (optional)
data = data.reset_index(drop=True)

data['streams'].isnull().sum()
```

```
0
```

```python
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 952 entries, 0 to 951
Data columns (total 7 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   track_name            952 non-null    object 
 1   artist(s)_name        952 non-null    object 
 2   artist_count          952 non-null    int64  
 3   released_year         952 non-null    int64  
 4   in_spotify_playlists  952 non-null    int64  
 5   in_spotify_charts     952 non-null    int64  
 6   streams               952 non-null    float64
dtypes: float64(1), int64(4), object(2)
memory usage: 52.2+ KB
```

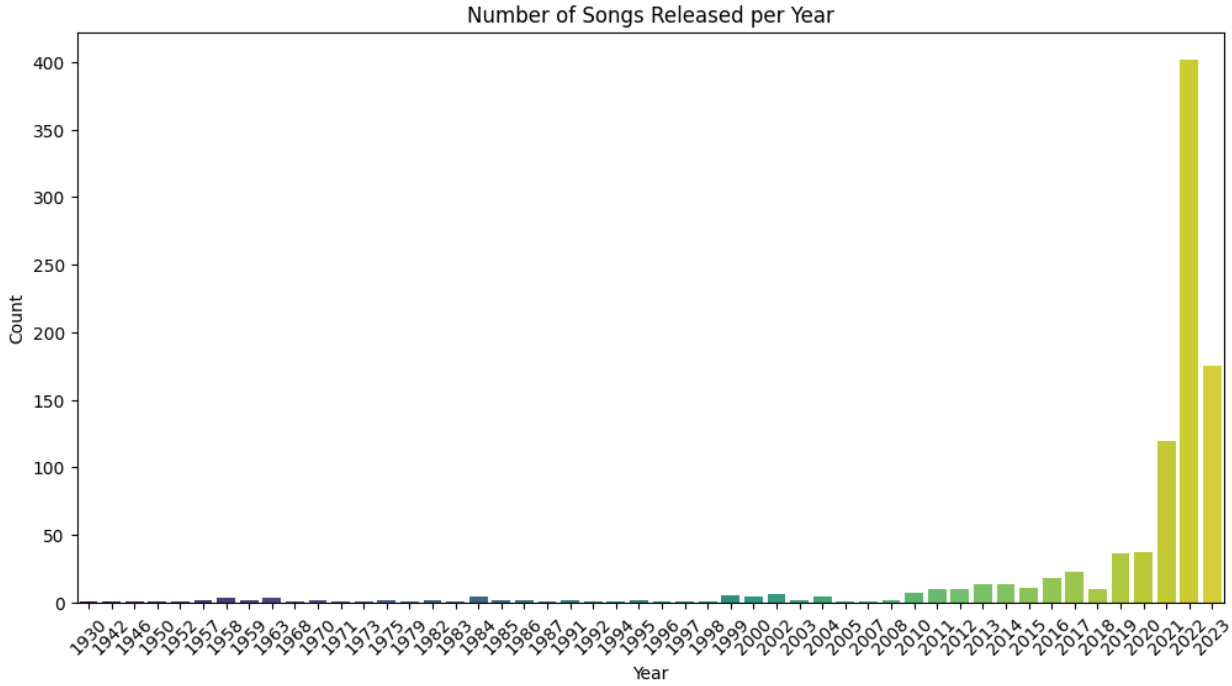
```python
df['artist(s)_name'].value_counts()
```

```
artist(s)_name
Taylor Swift                 34
The Weeknd                   22
Bad Bunny                    19
SZA                          19
Harry Styles                 17
                             ..
Karol G, Ovy On The Drums     1
Coolio, L.V.                  1
Kordhell                      1
Kenia OS                      1
Feid, Sech, Jhayco            1
Name: count, Length: 645, dtype: int64
```

```python
df['released_year'].unique()
```

```
array([2023, 2019, 2022, 2013, 2014, 2018, 2017, 2020, 2016, 2012,
1999,
       2008, 1975, 2021, 2015, 2011, 2004, 1985, 2007, 2002, 2010,
1983,
       1992, 1968, 1984, 2000, 1997, 1995, 2003, 1973, 1930, 1994,
1958,
       1957, 1963, 1959, 1970, 1971, 1952, 1946, 1979, 1950, 1942,
1986,
       2005, 1991, 1996, 1998, 1982, 1987])
```

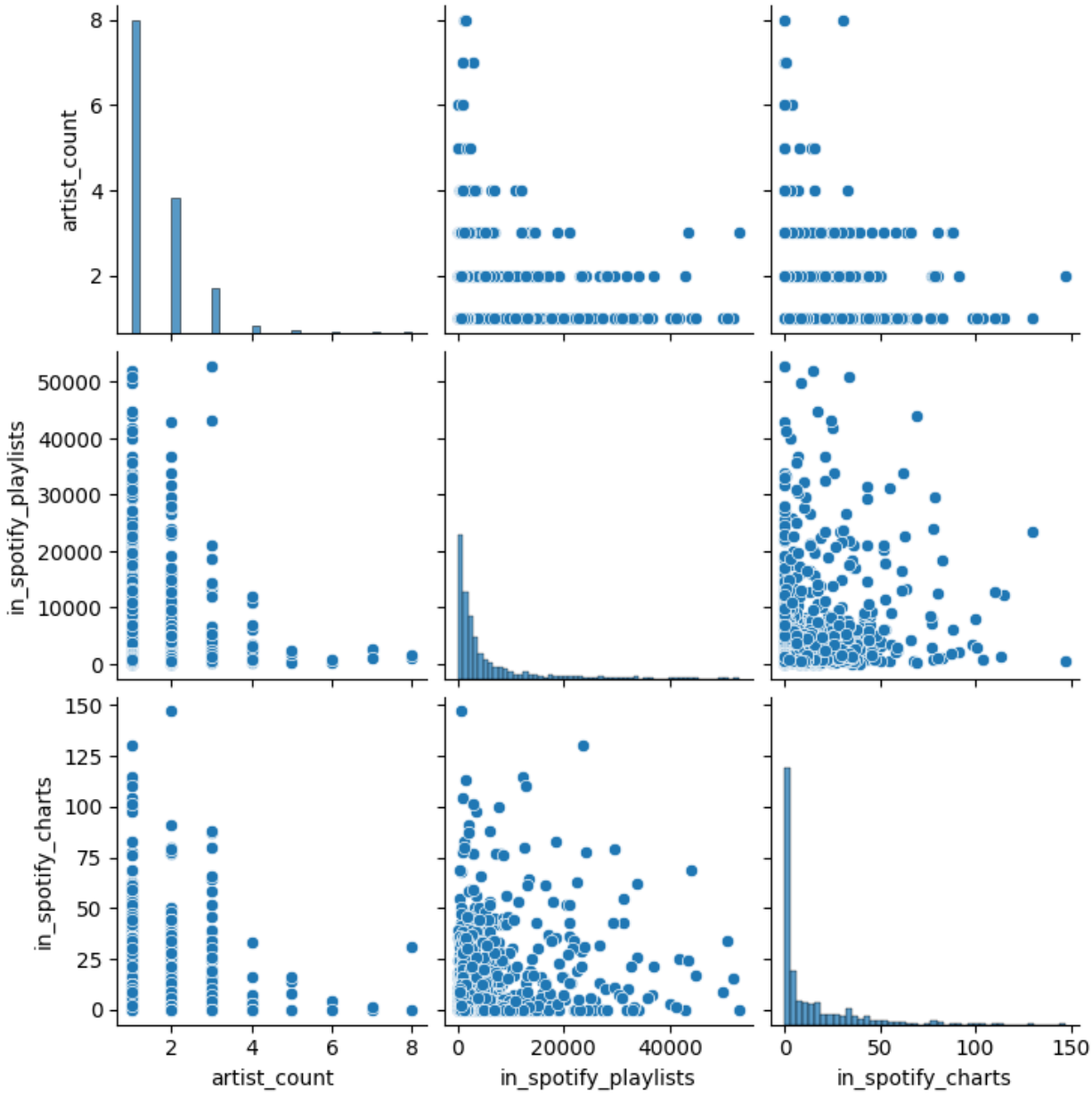
```
plt.figure(figsize=(12, 6))
sns.countplot(data=df, x='released_year', palette='viridis')
plt.xlabel('Year')
plt.ylabel('Count')
plt.title('Number of Songs Released per Year')
plt.xticks(rotation=45)
plt.show()
```



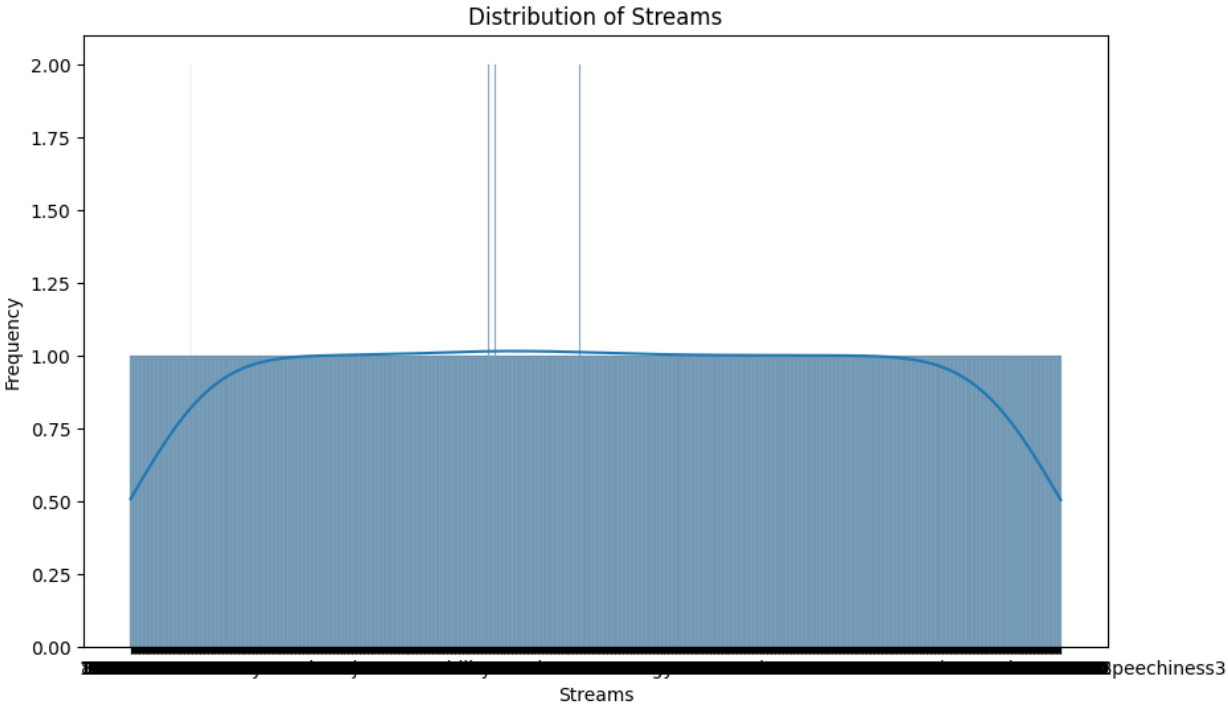

```
sns.pairplot(df[['artist_count', 'in_spotify_playlists',
'in_spotify_charts', 'streams']])
```

```
/home/rohit/.local/lib/python3.10/site-packages/seaborn/
axisgrid.py:118: UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```

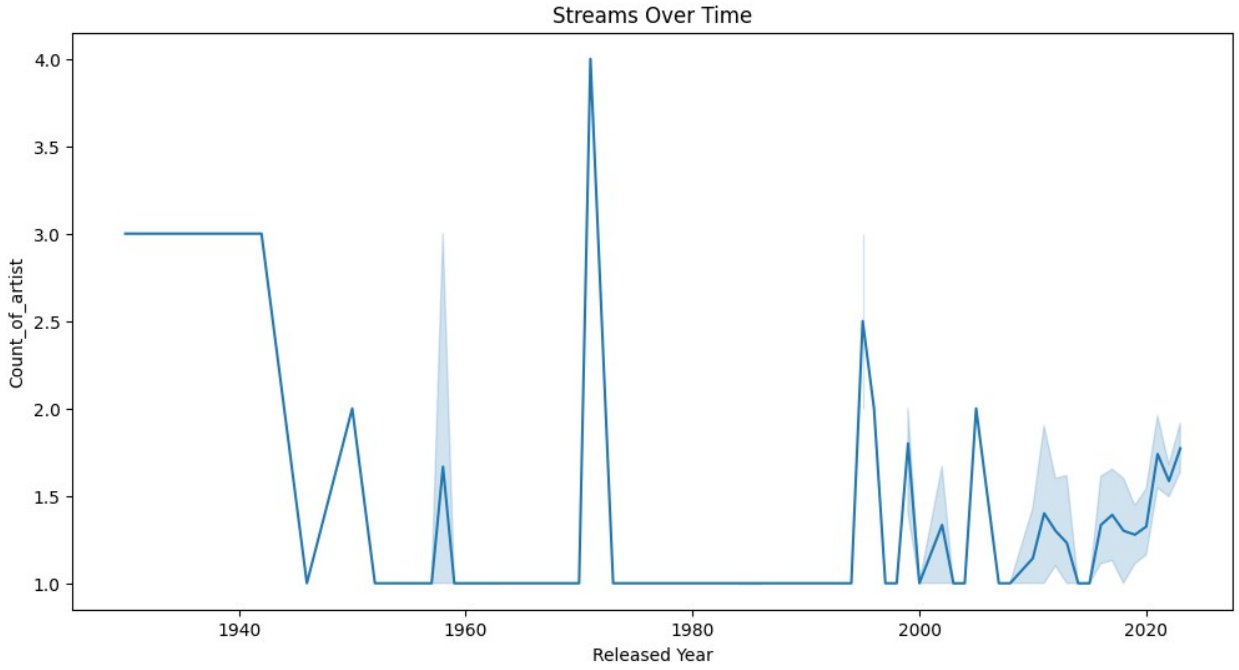
```
<seaborn.axisgrid.PairGrid at 0x7efd19b439d0>
```



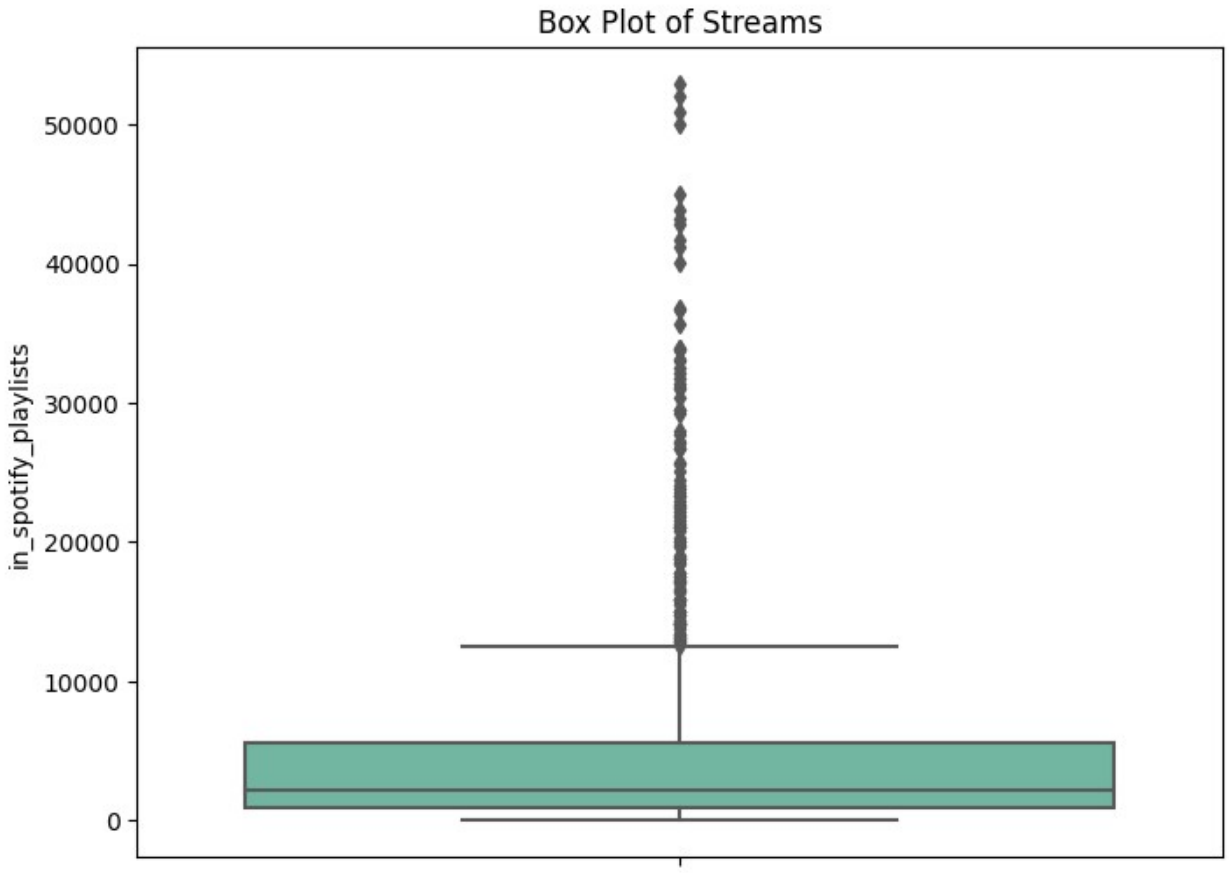

```
plt.figure(figsize=(10, 6))
sns.histplot(data=df, x='streams', bins=30, kde=True)
plt.xlabel('Streams')
plt.ylabel('Frequency')
plt.title('Distribution of Streams')
plt.show()
```




```
plt.figure(figsize=(12, 6))
sns.lineplot(data=df, x='released_year', y='artist_count')
plt.xlabel('Released Year')
plt.ylabel('Count_of_artist')
plt.title('Streams Over Time')
plt.show()
```




```
plt.figure(figsize=(8, 6))
sns.boxplot(data=df, y='in_spotify_playlists', palette='Set2')
plt.title('Box Plot of Streams')
plt.show()
```


Box Plot of Streams
in_spotify_playlists
50000
40000
30000
20000
10000
0