Cyrillic, Stones Preussia & Beatles

160 000+ tracks on Spotify

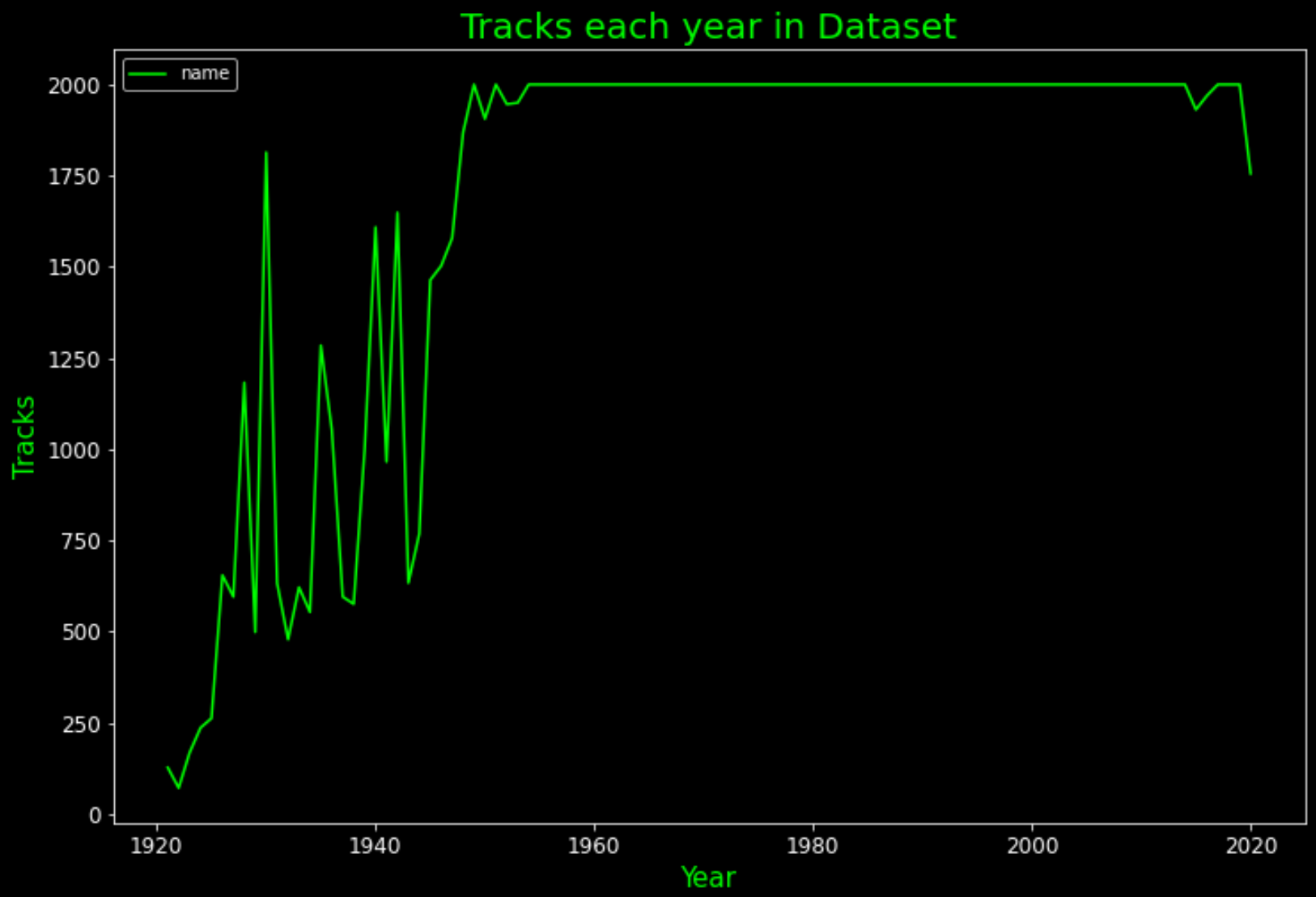
We'll look at

- some overall stats
- tendencies over the years
- most popular artists
- Stones & Beatles
- the Typical ****

The Data
169909 tracks
1921 – 2020

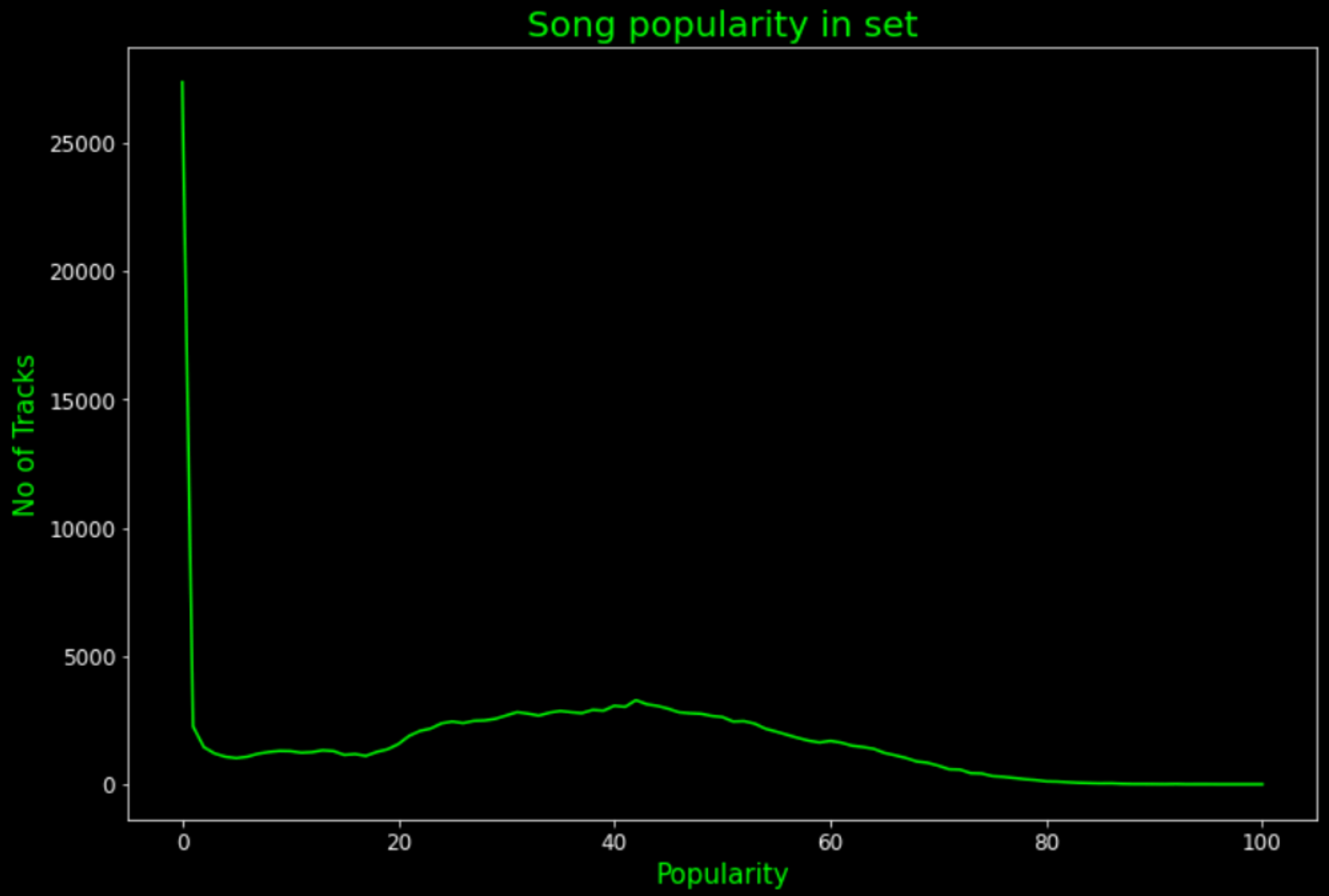
Data set

- Artist
- Acousticness – value between 0&1
- Danceability – value between 0&1
- Duration_ms
- Energy – value between 0&1
- Explicit – binary
- Id
- Instrumentalness – value between 0&1
- Key- main key as int 0-11
- Liveness – value between 0&1
- Loudness – (-60)-0 db
- Mode – binary
- Name
- Popularity – value between 0&1
- Release date
- Speechiness – value between 0&1
- Tempo – bpm
- Valence – value between 0&1
- Year

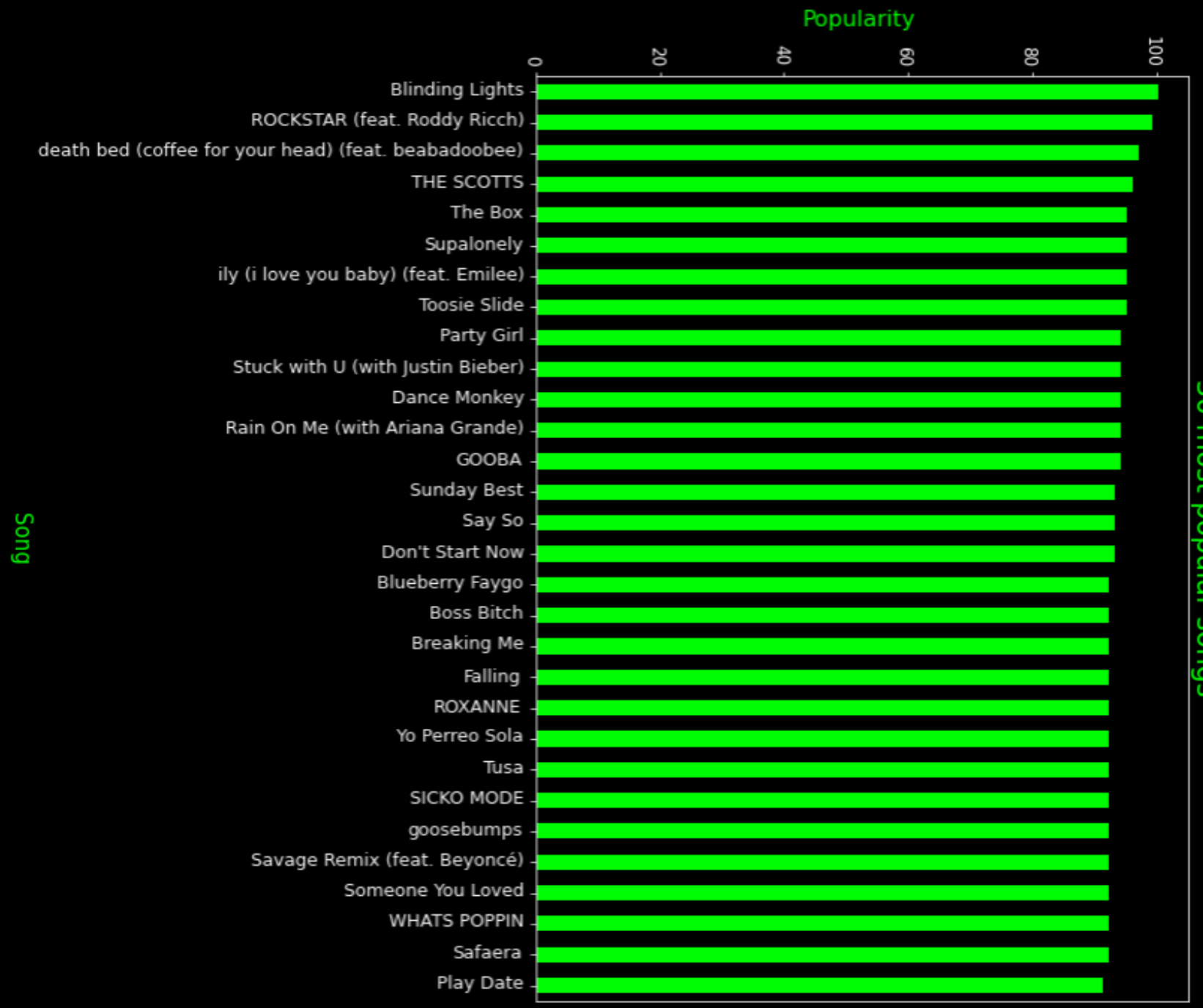


Number of songs from each year that are present in the dataset

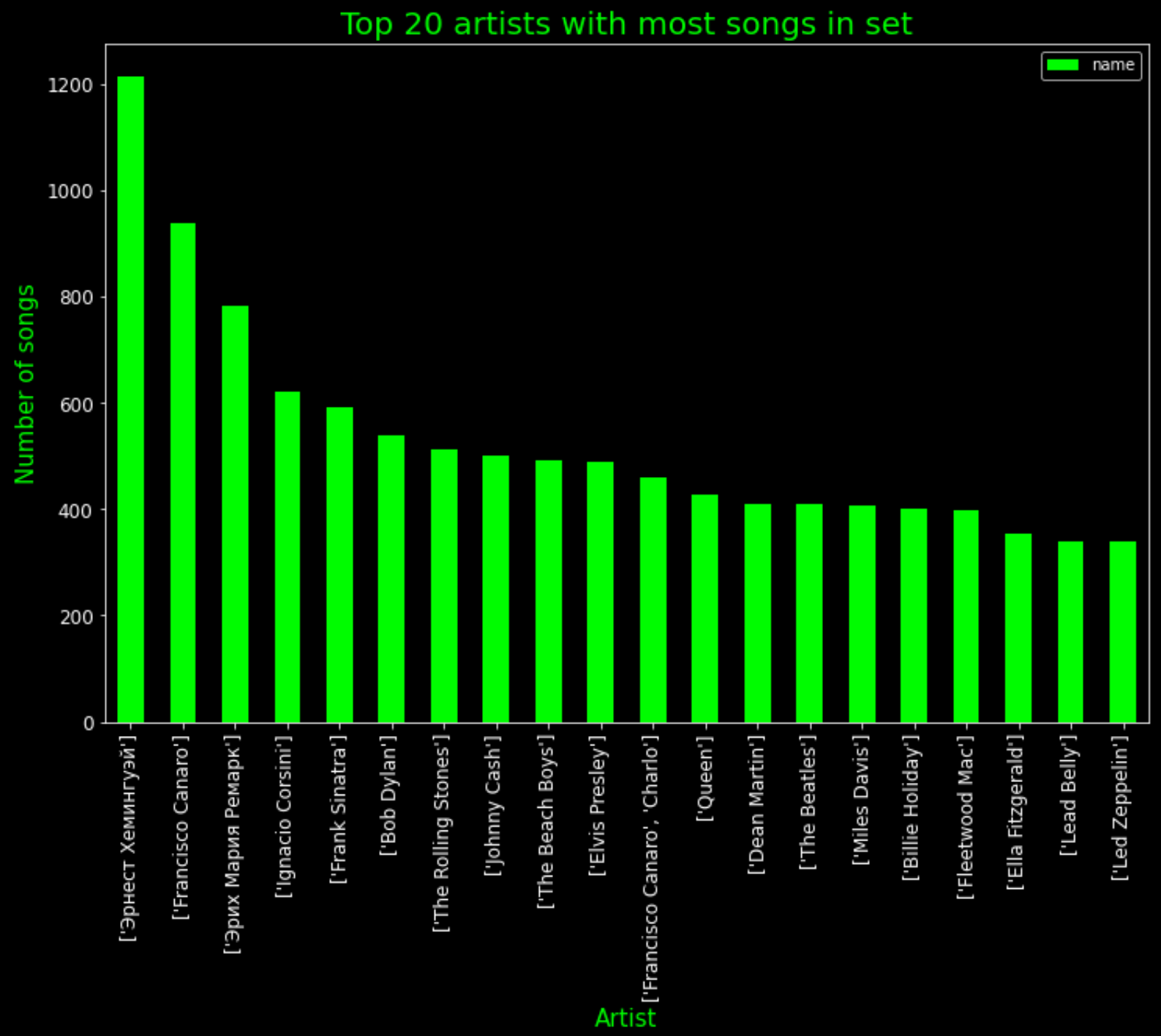
Distribution of popularity.
A popularity of 0 most common.



30 most popular songs.
These are all from between 2015 & 2020

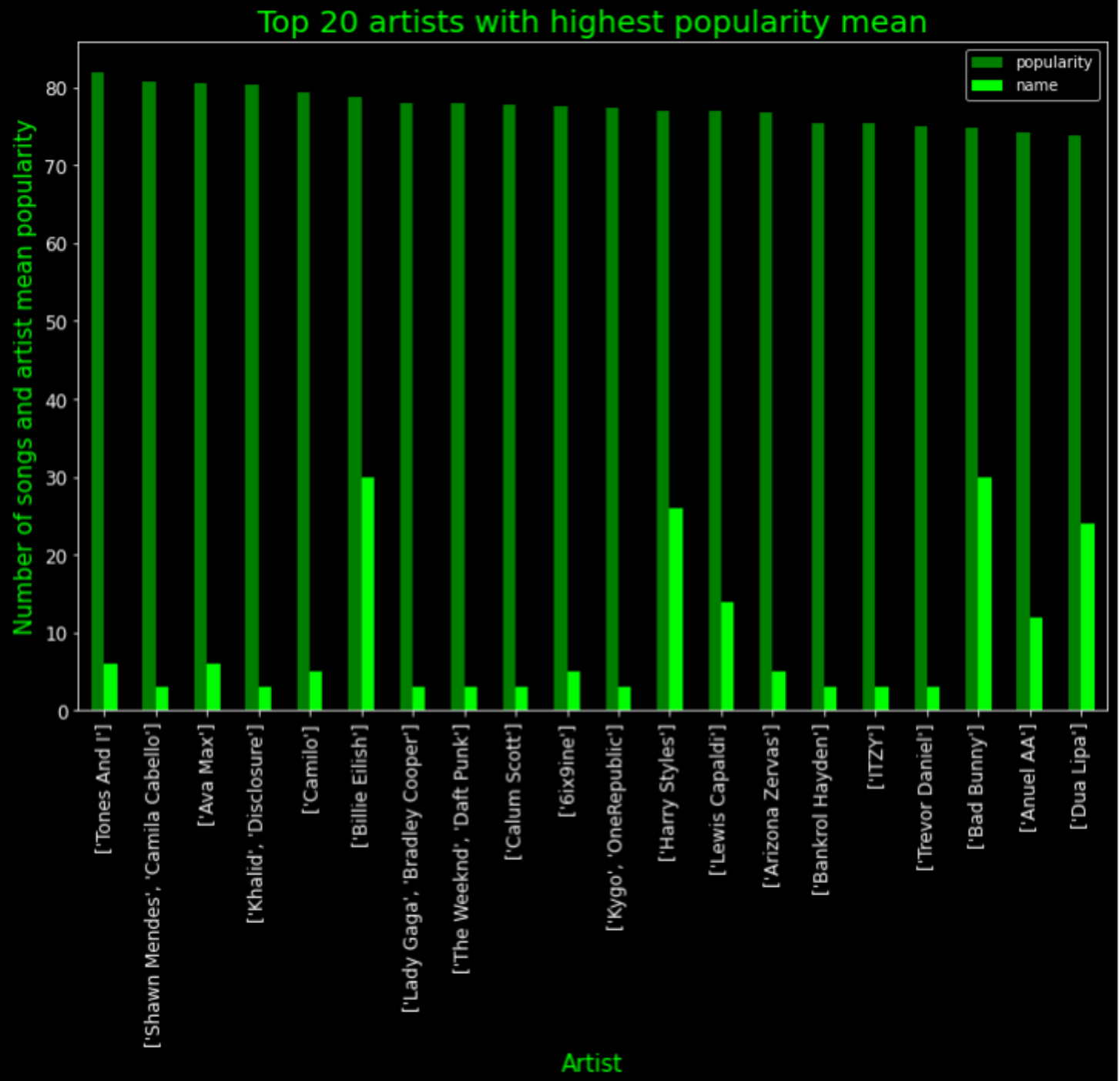


The 20 artists with the most tracks in the dataset



The 20 artists with the highest mean popularity for all their songs.
Only looking at artists with 2 or more songs to exclude featurings and one-hit-wonders.

How many songs they have in the dataset is also presented in the chart.



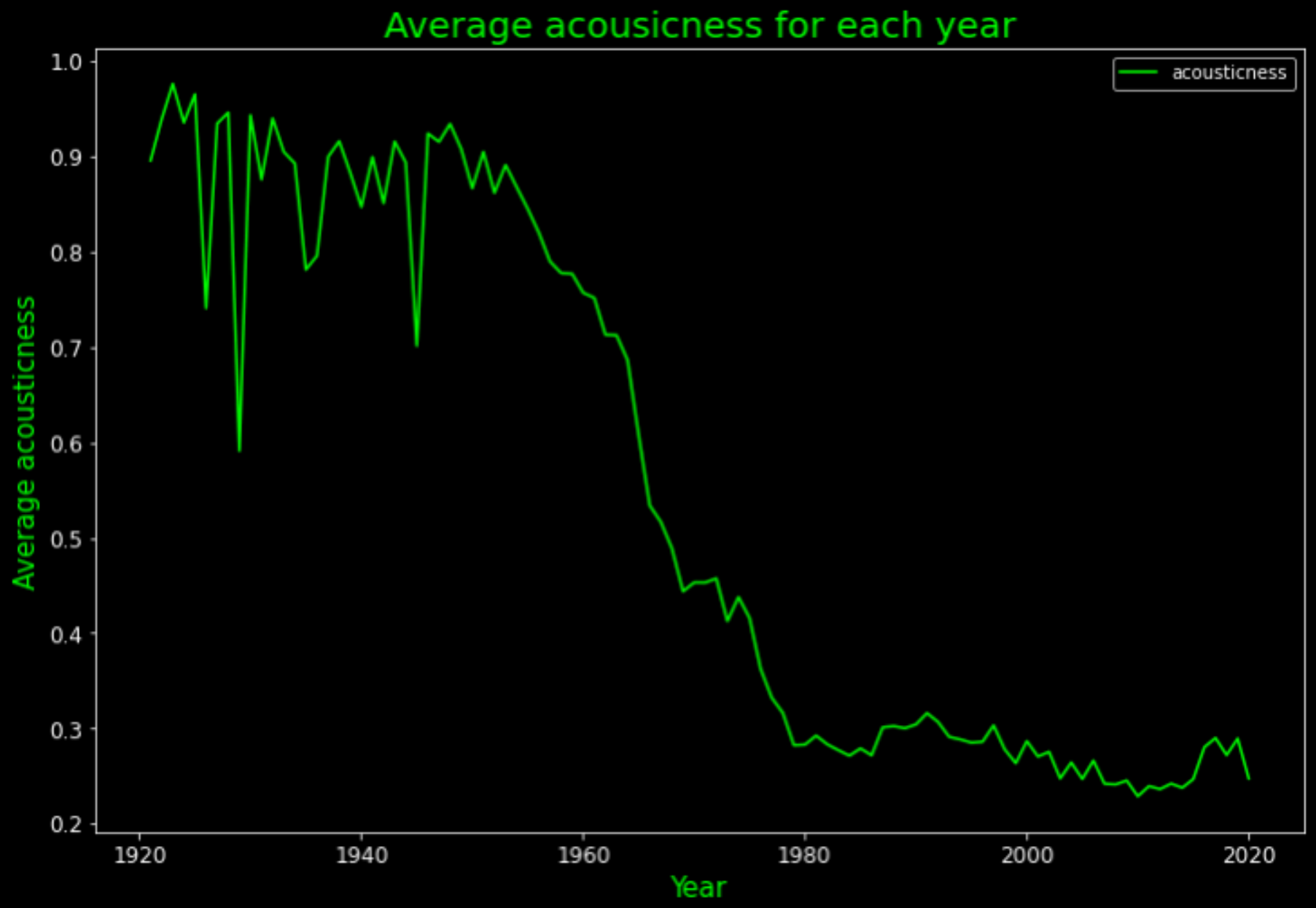
Over the years

Let’s take a look at whether there have been any trends over the years in terms of

- Mean acousticness
- Percentage of explicit songs
- Mean popularity
- Mean valence & BPM
- Mean danceability & energy

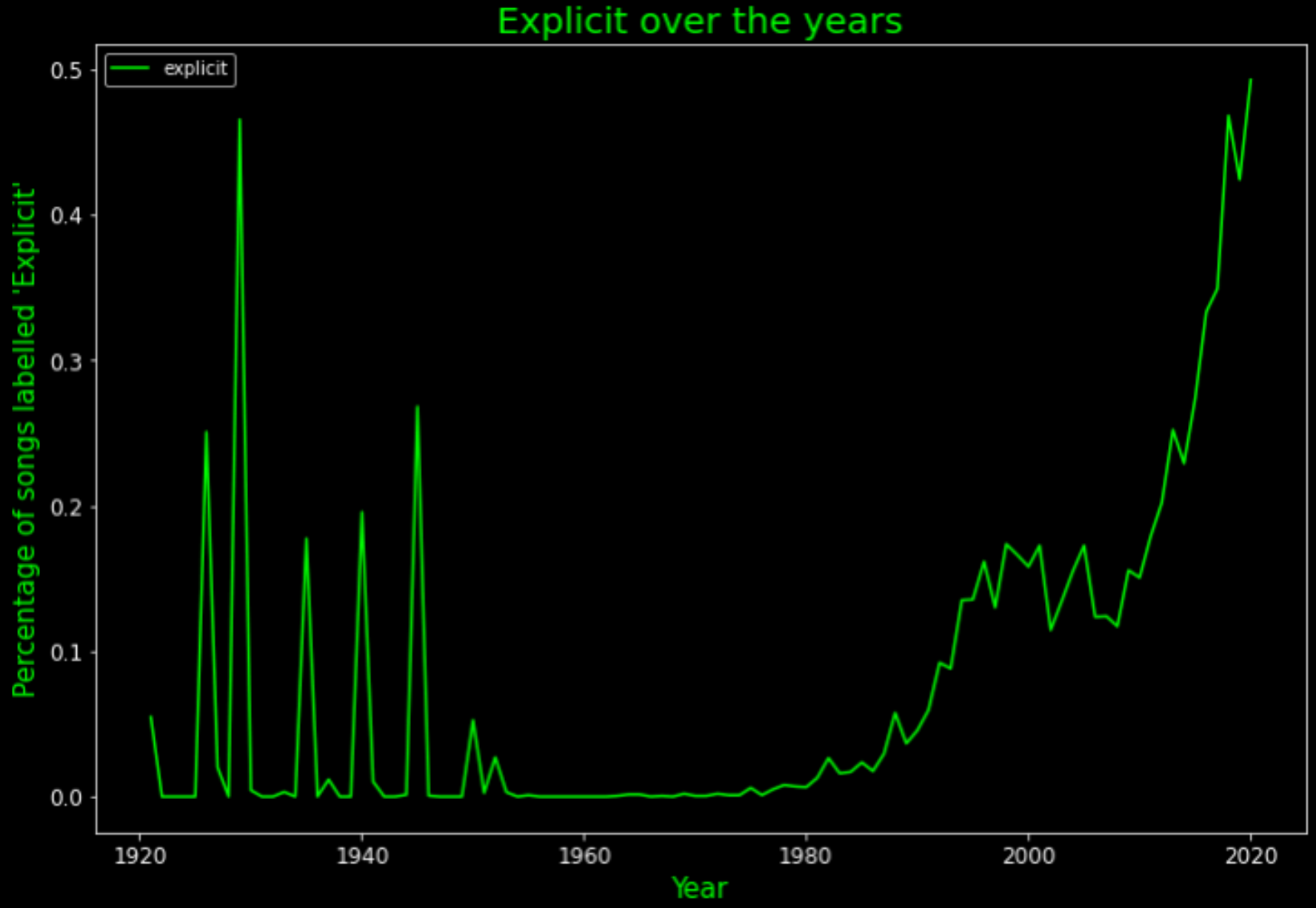
Acousticness

As you can see there are a few peaks before and around 1940 that look a bit suspicious



Explicit

When looking at the percentage of songs for each year labelled explicit those peaks appear again



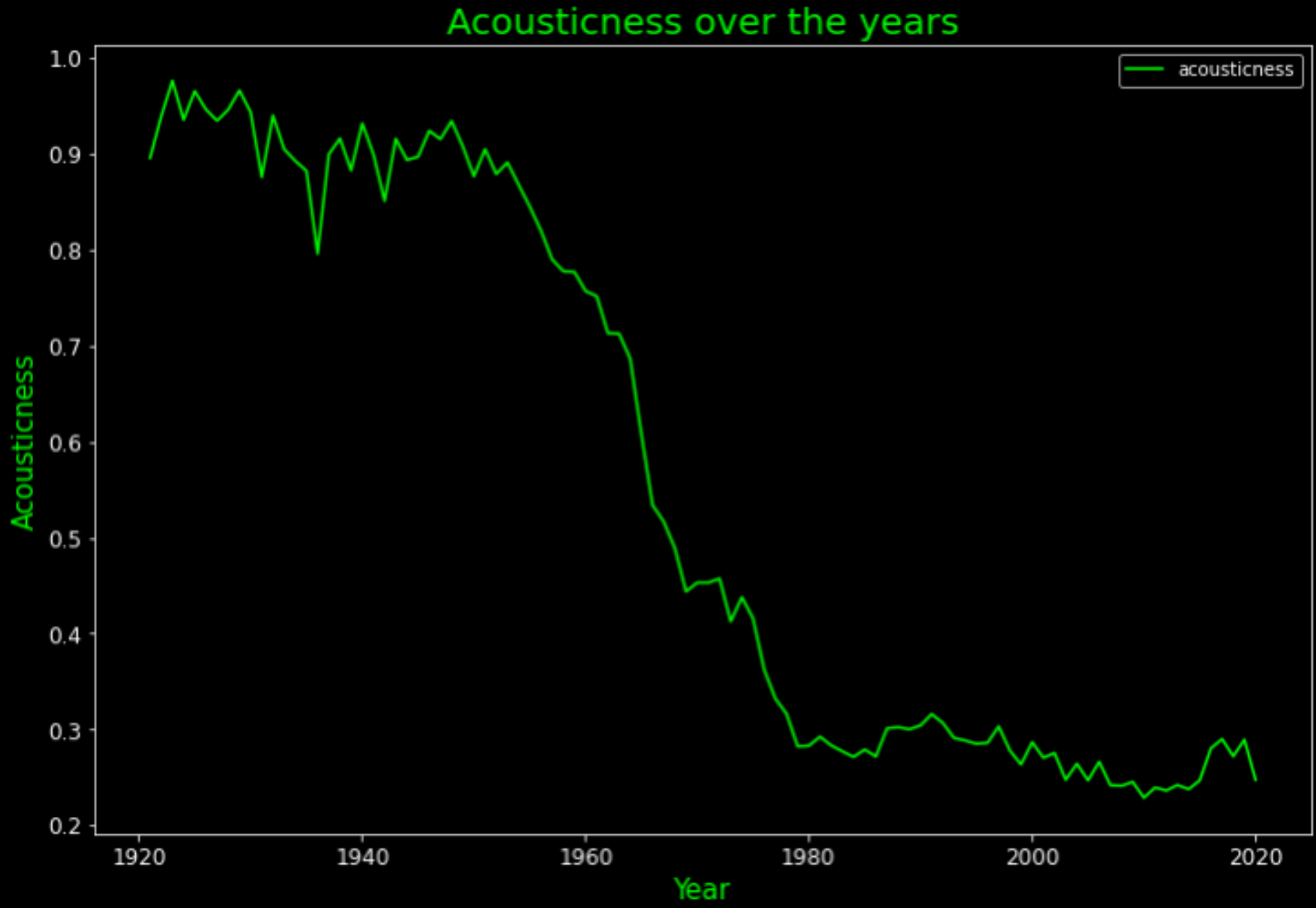
Remember these two?

- Эрих Мария Ремарк
- Эрнест Хемингуэй
  - All of their songs are explicit
  - They are the most and third most represented artists in the dataset

Hence, we got rid of them

Acousticness updated

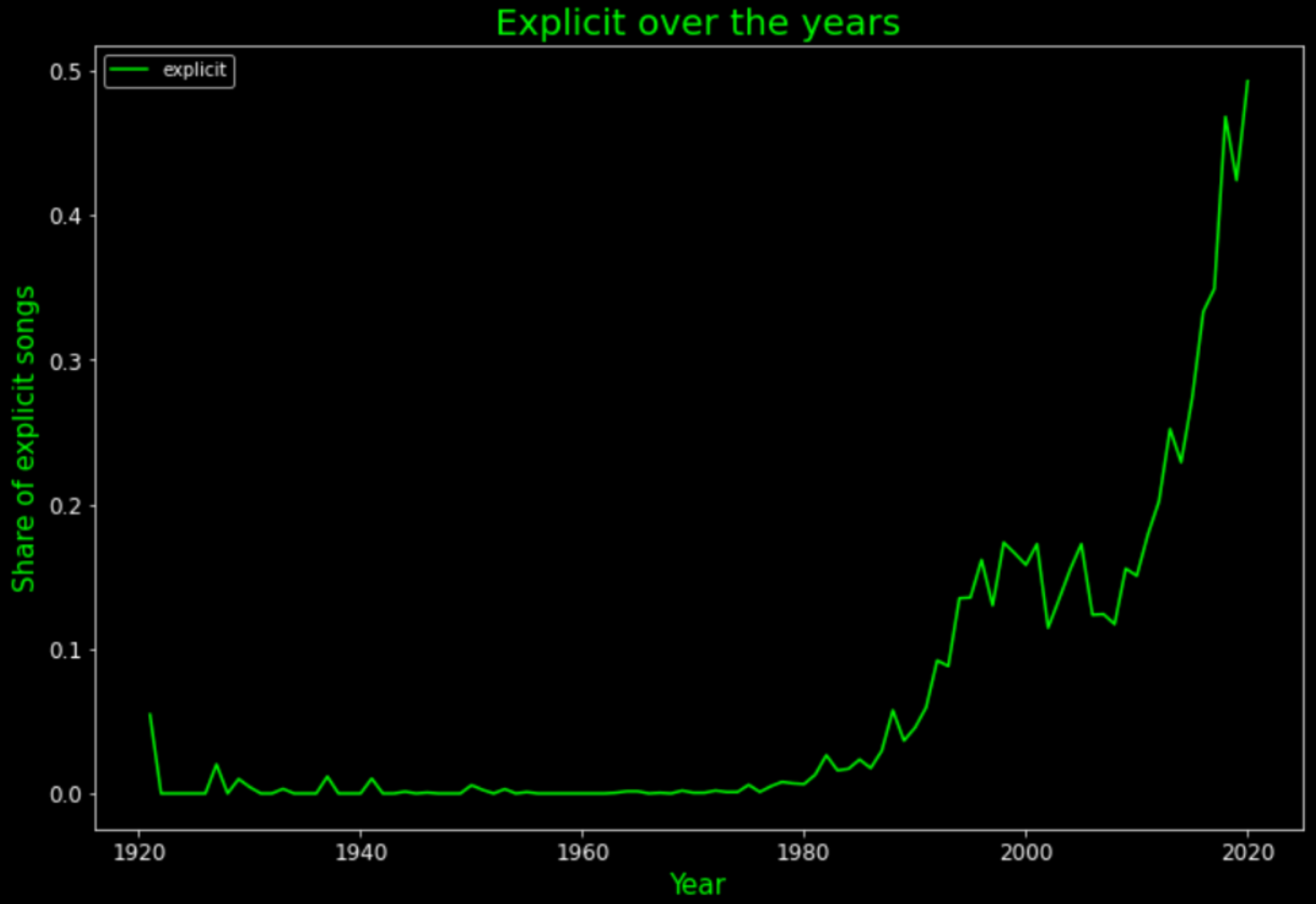
An updated graph over mean acousticness over the years.
There is a negative correlation of 0.94



Explicit updated

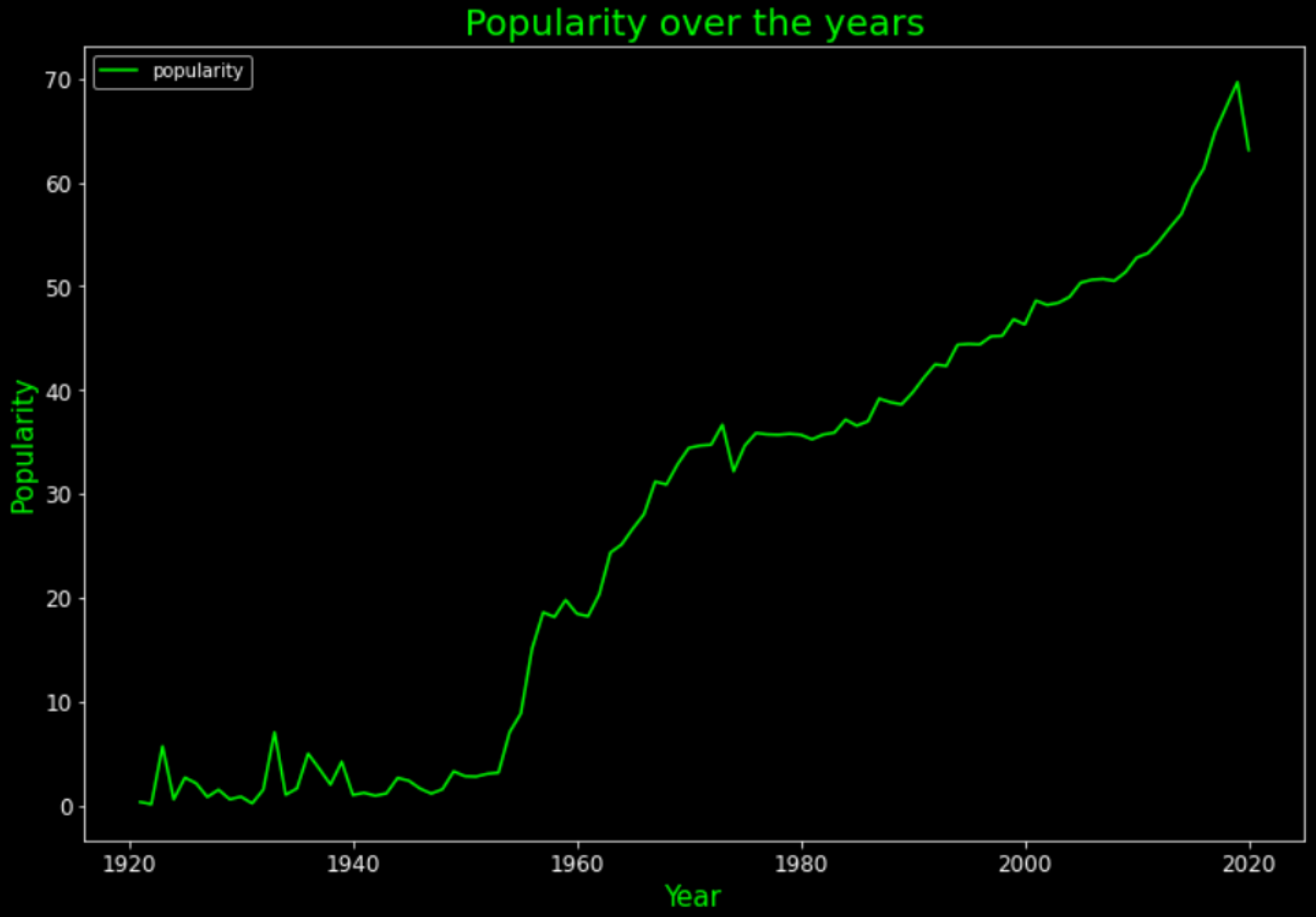
An updated graph over percentage of explicit tracks over the years.
There is a correlation of 0.75

Explicit and acousticness have a correlation of -0.6



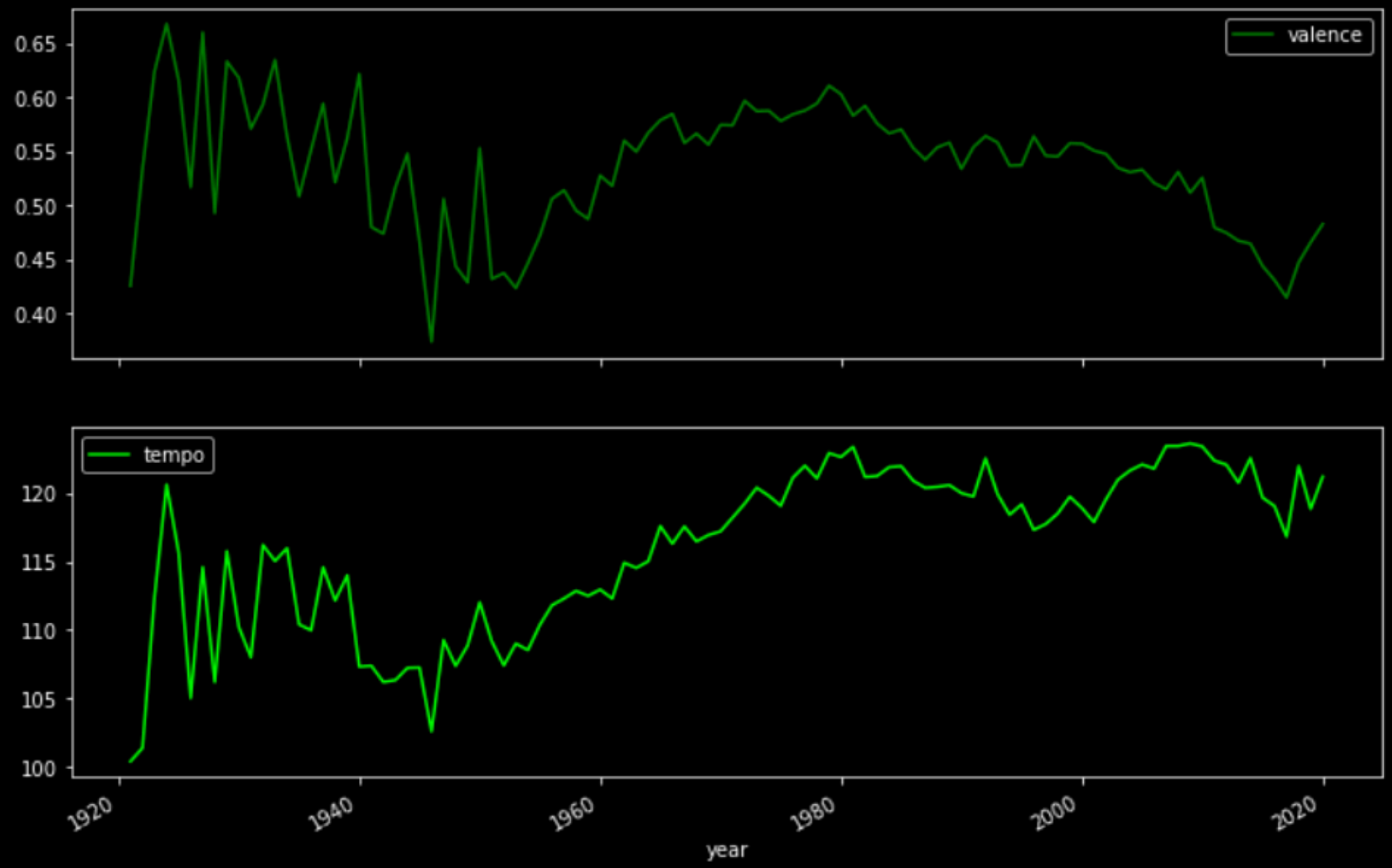
Popularity

- Popularity is biased
- There is a drop in popularity 2020
- In 2020 there are 78 tracks with popularity of 0
- All of those were released right before the dataset was created



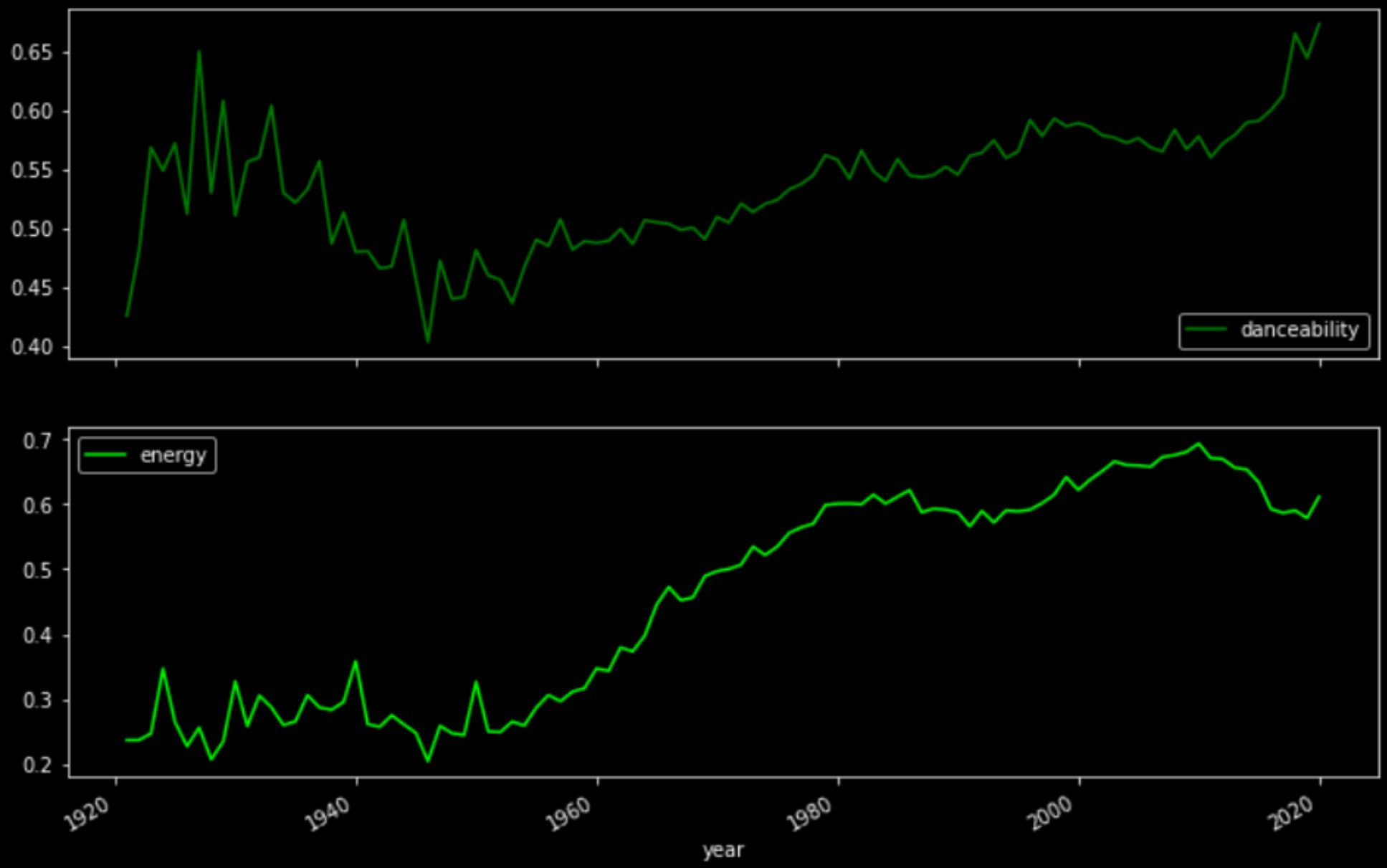
Valence & Tempo

Mean valence & tempo over the years



Danceability & Energy

Mean danceability & energy over the years



The top 10 %

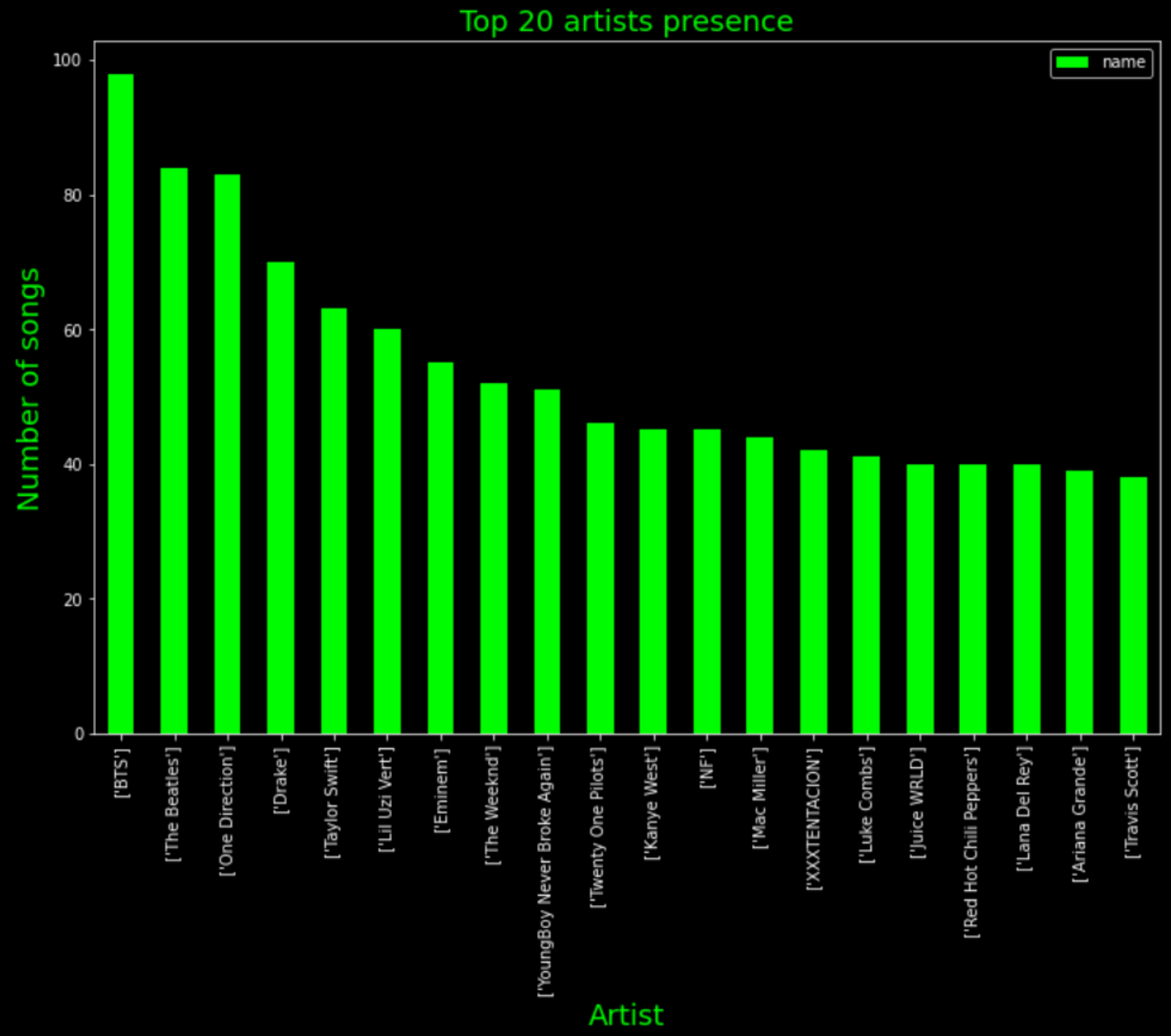
Sorted out the 10% most popular tracks

In those 10% - which artists have the most songs?

In those 10% - which artists have the highest mean popularity?

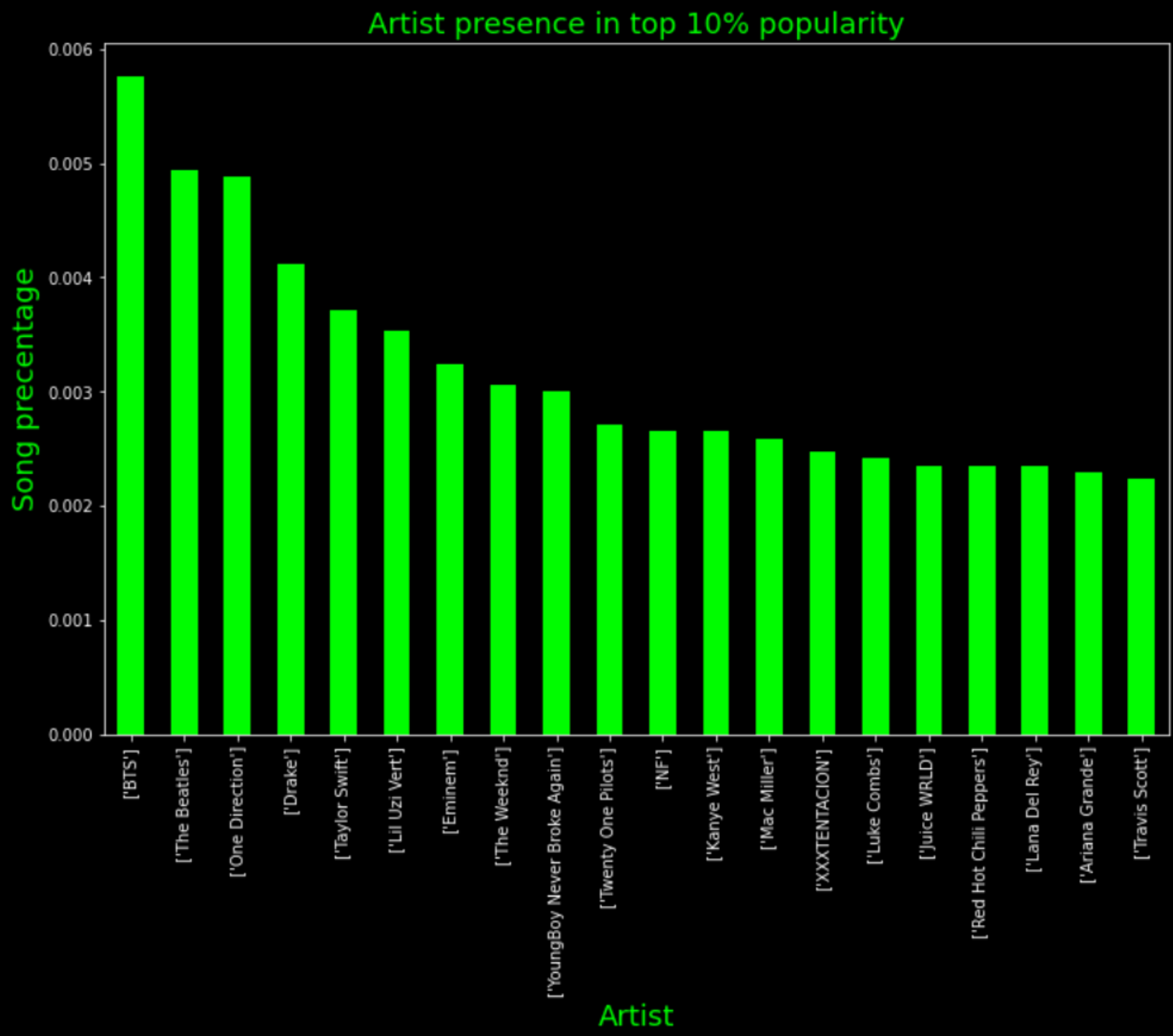
Among the top 10% most popular tracks,
which artists hold the highest presence?

Here with artists' presence as number of songs



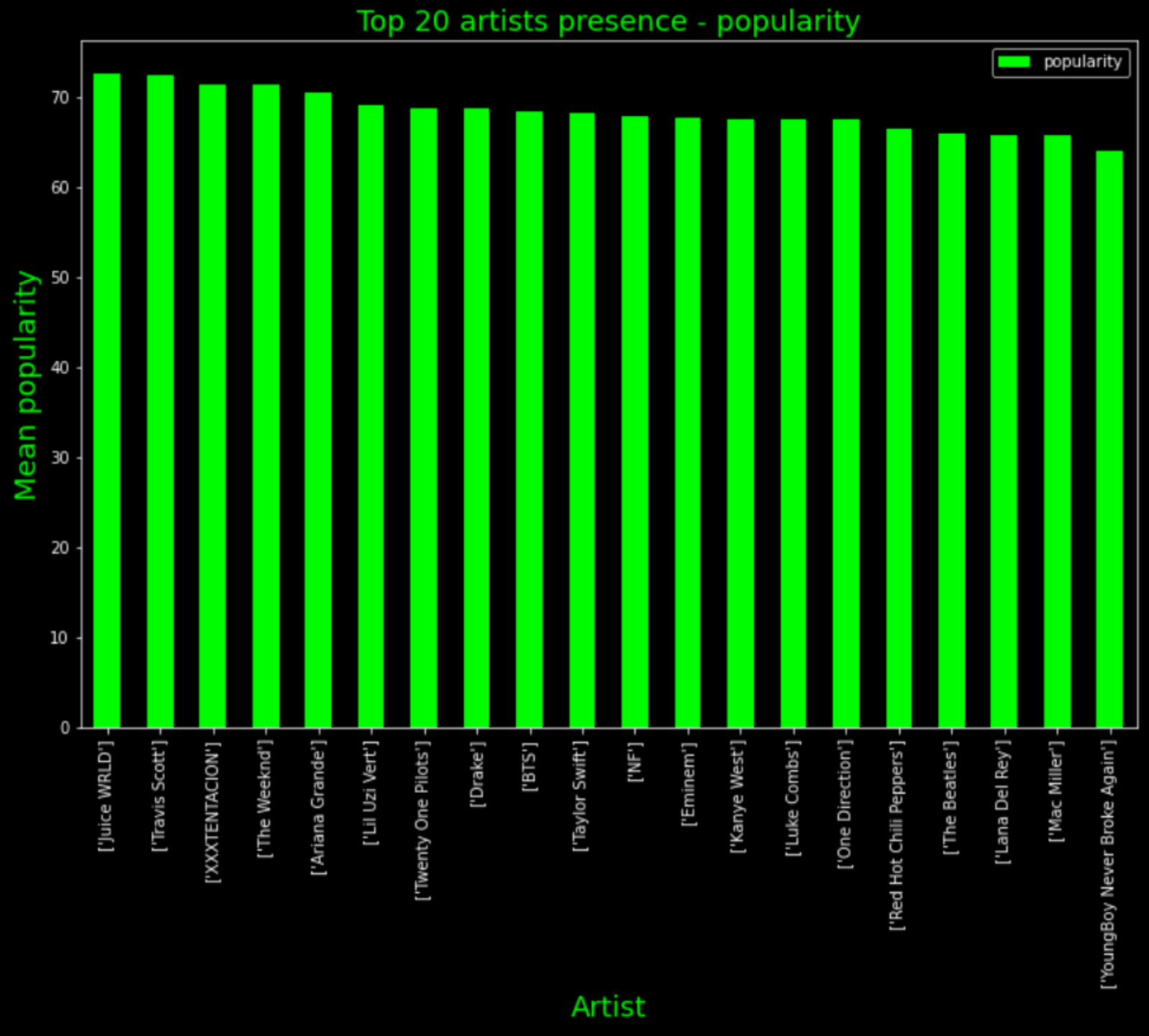
Among the top 10% most popular tracks,
which artists hold the highest presence?

Here with artists' presence as percentage

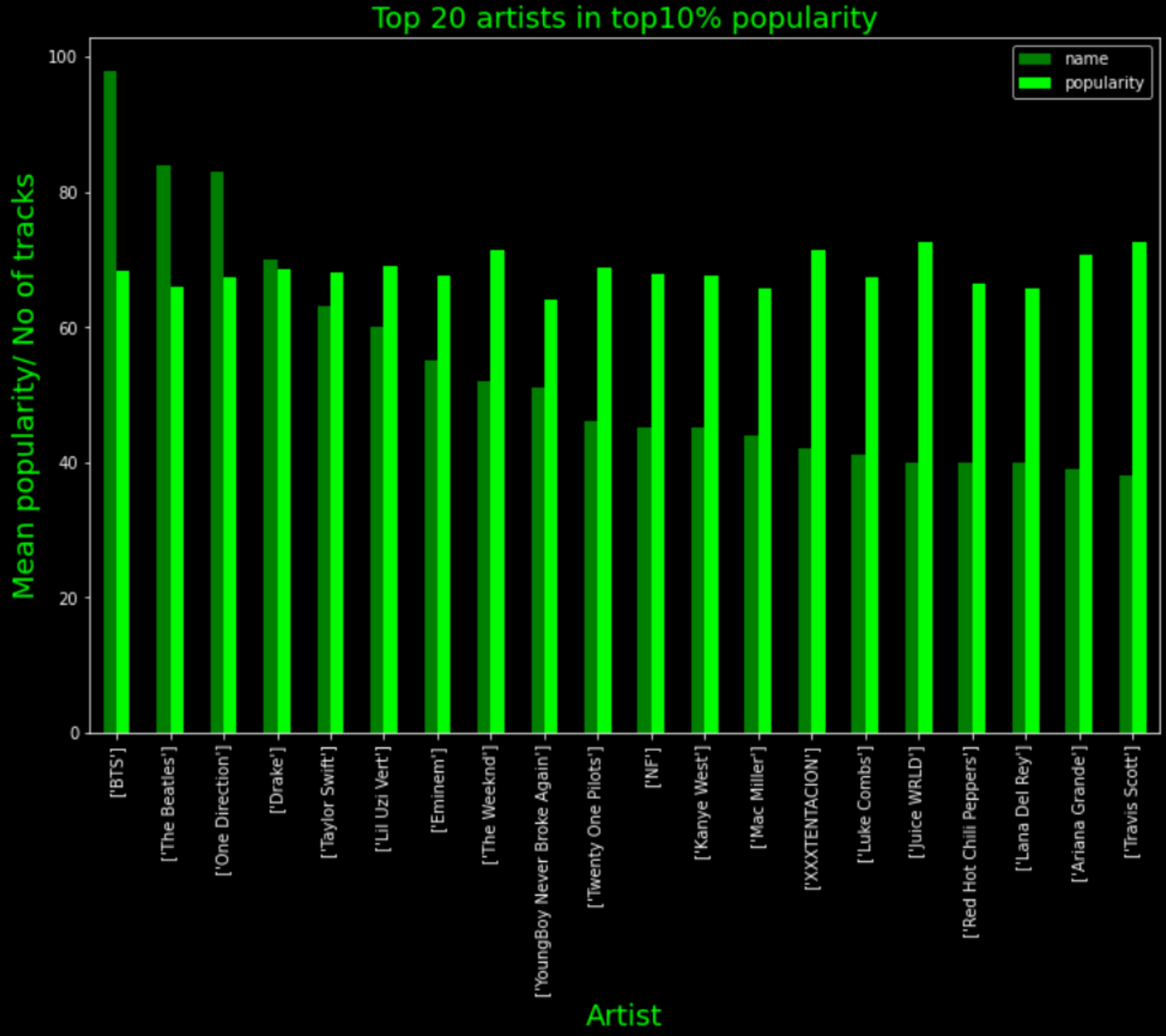


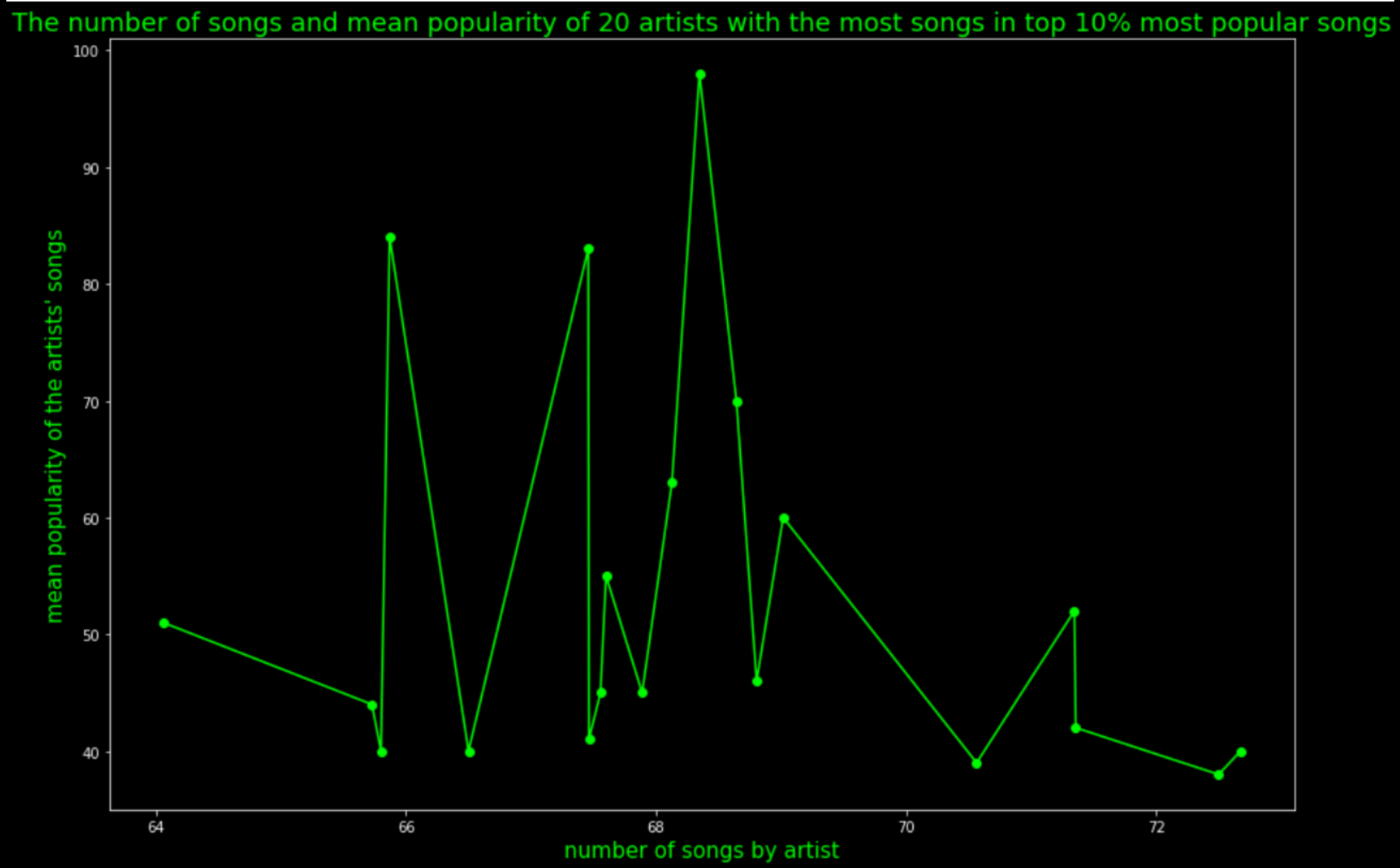
Among the top 10% most popular tracks,
which artists hold the highest presence?

Here with the artists' mean popularity



The two previous bar charts presented next to each other

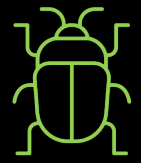



The number of songs and mean popularity of 20 artists with the most songs in top 10% most popular songs
mean popularity of the artists' songs
number of songs by artist
100
90
80
70
60
50
40
64
66
68
70
72

Beatles or Stones

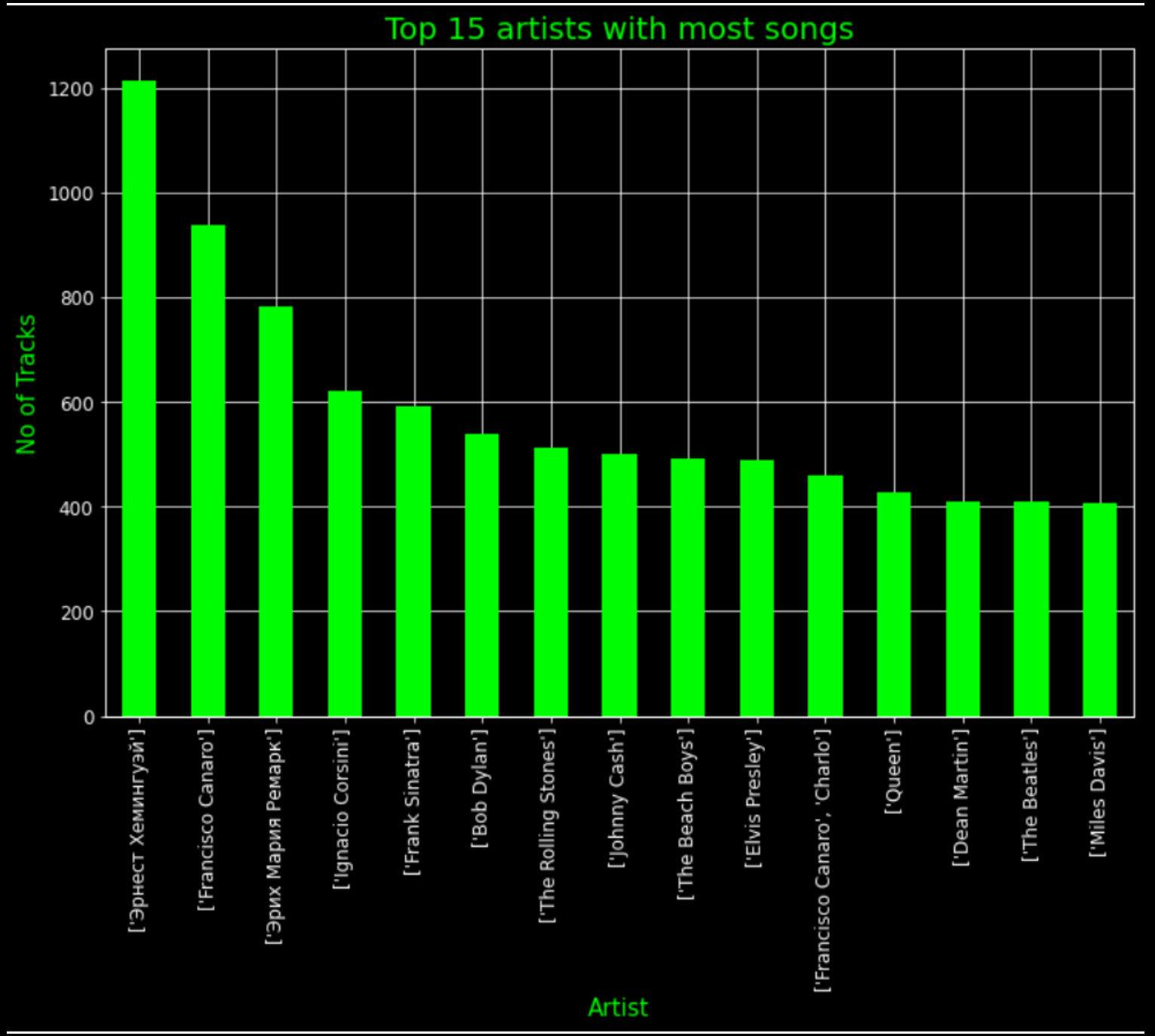
Let the world* decide

*(or at least this data set)

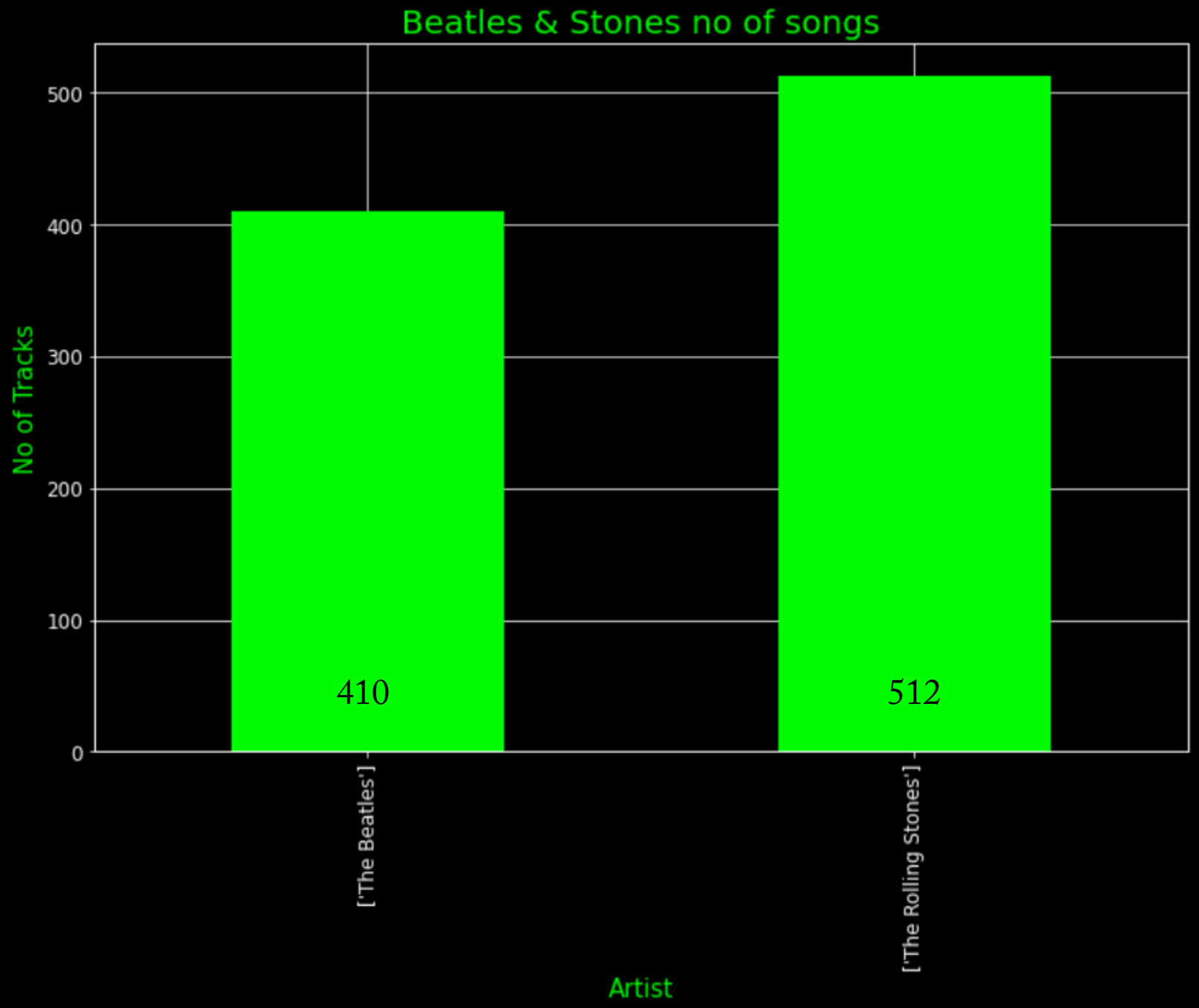


Stones have 7th most tracks in the set

Beatles have 14th most tracks in the set



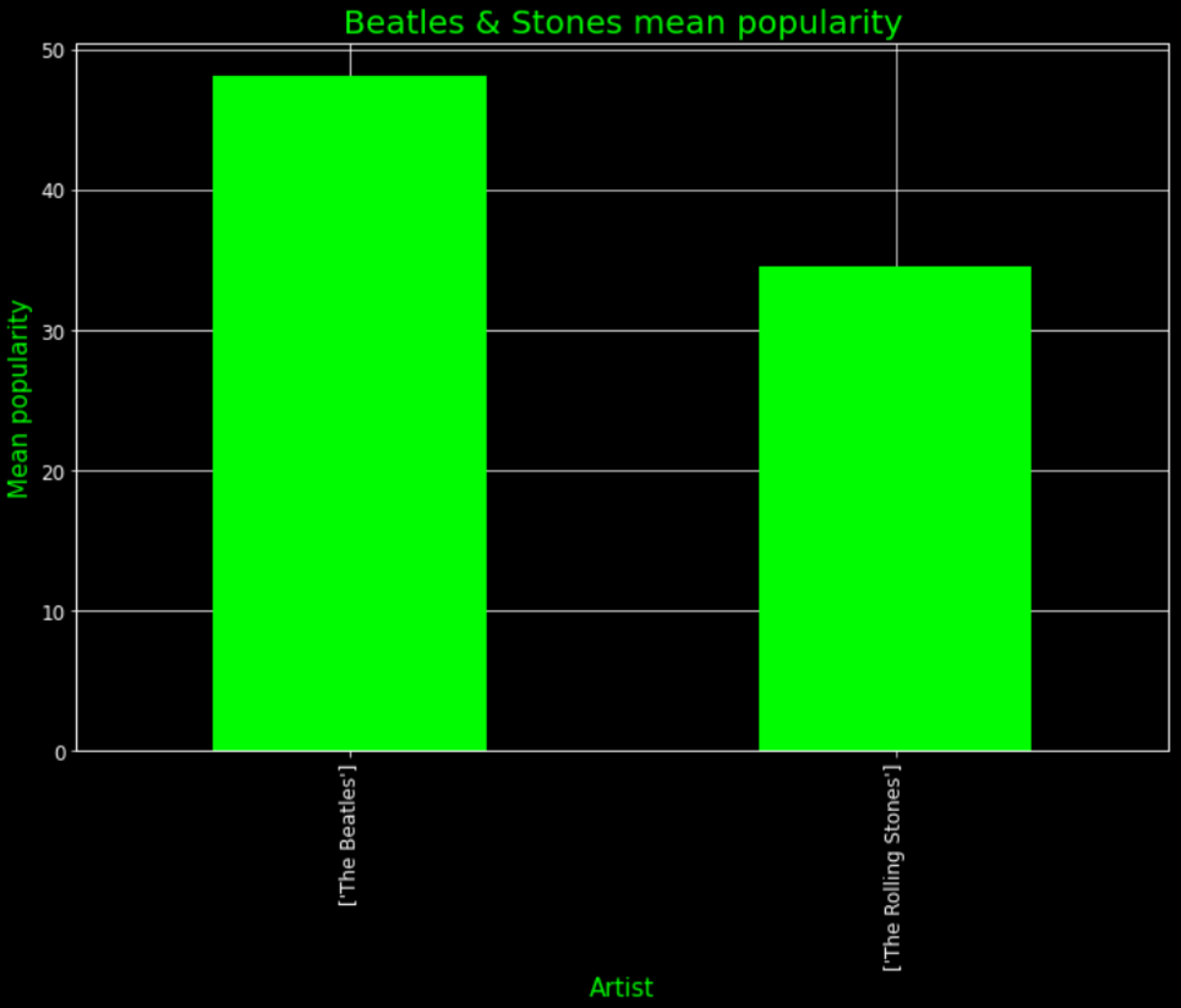
Looking only at the Beatles & Stones number of tracks



Artists' mean popularity:

Beatles: 48

Stones: 34,5

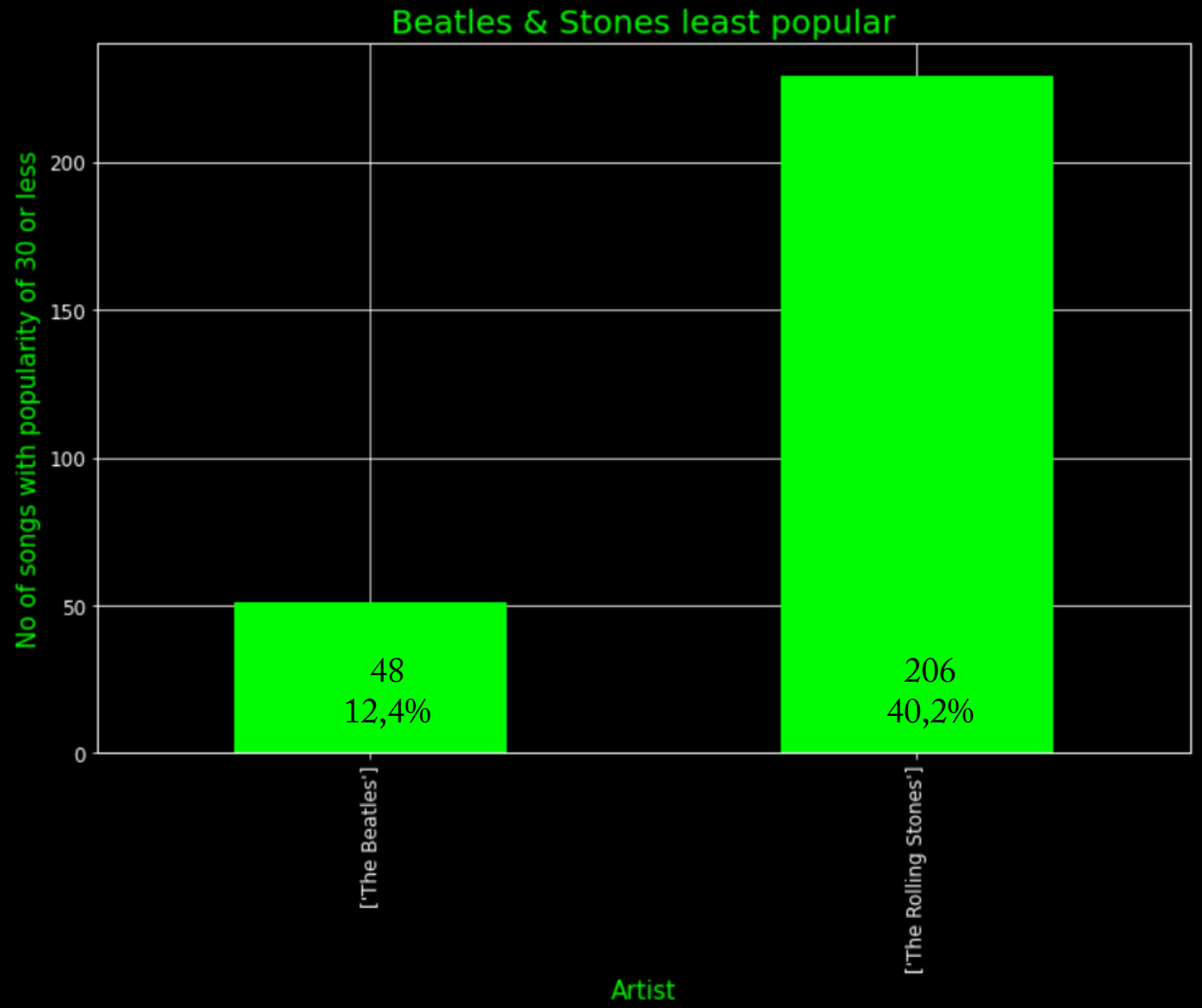


Percentage of songs with popularity of 75 or higher:

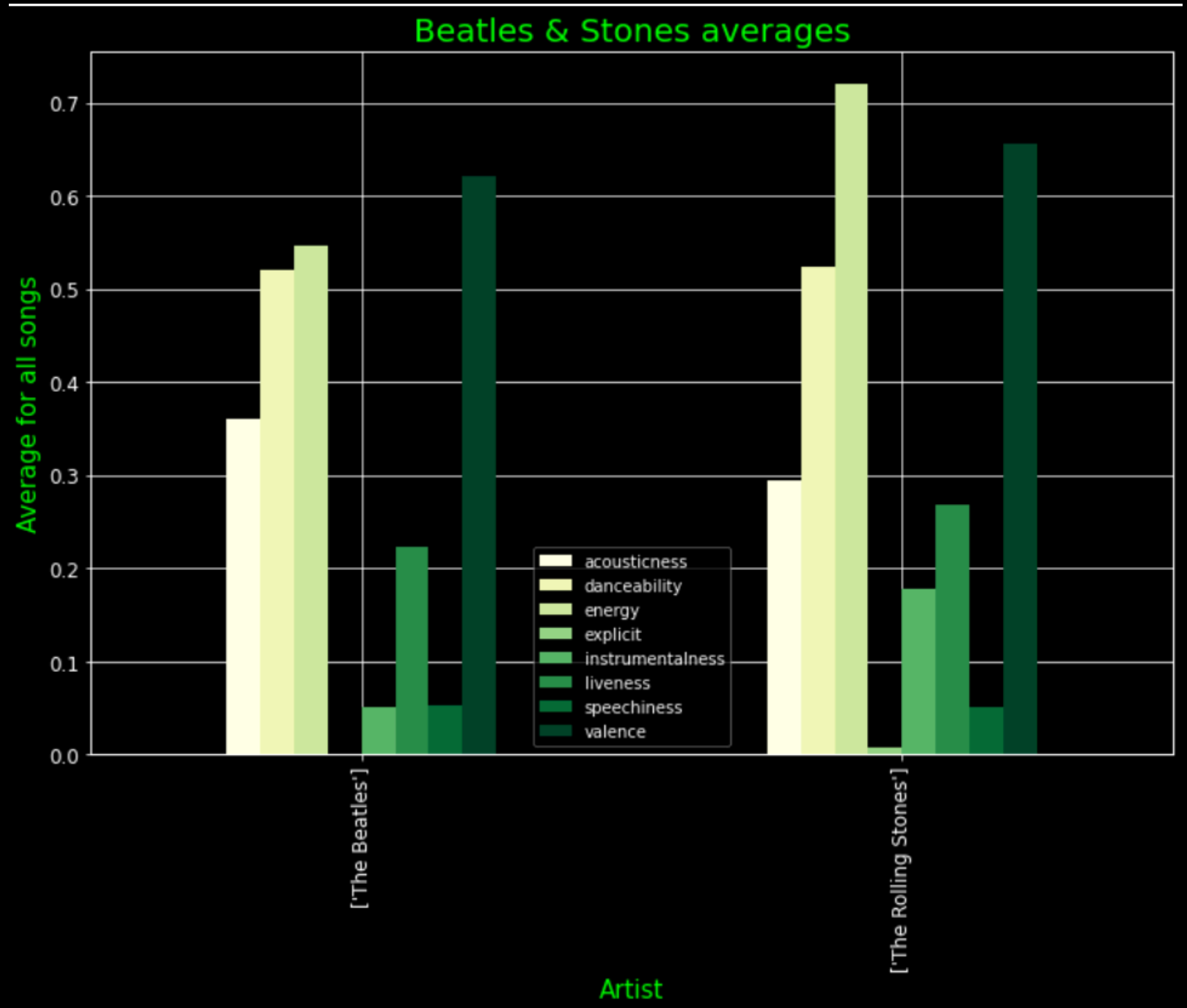
Beatles: 1,46%

Stones: 0,98%

Number of songs that have a popularity score of less than 30

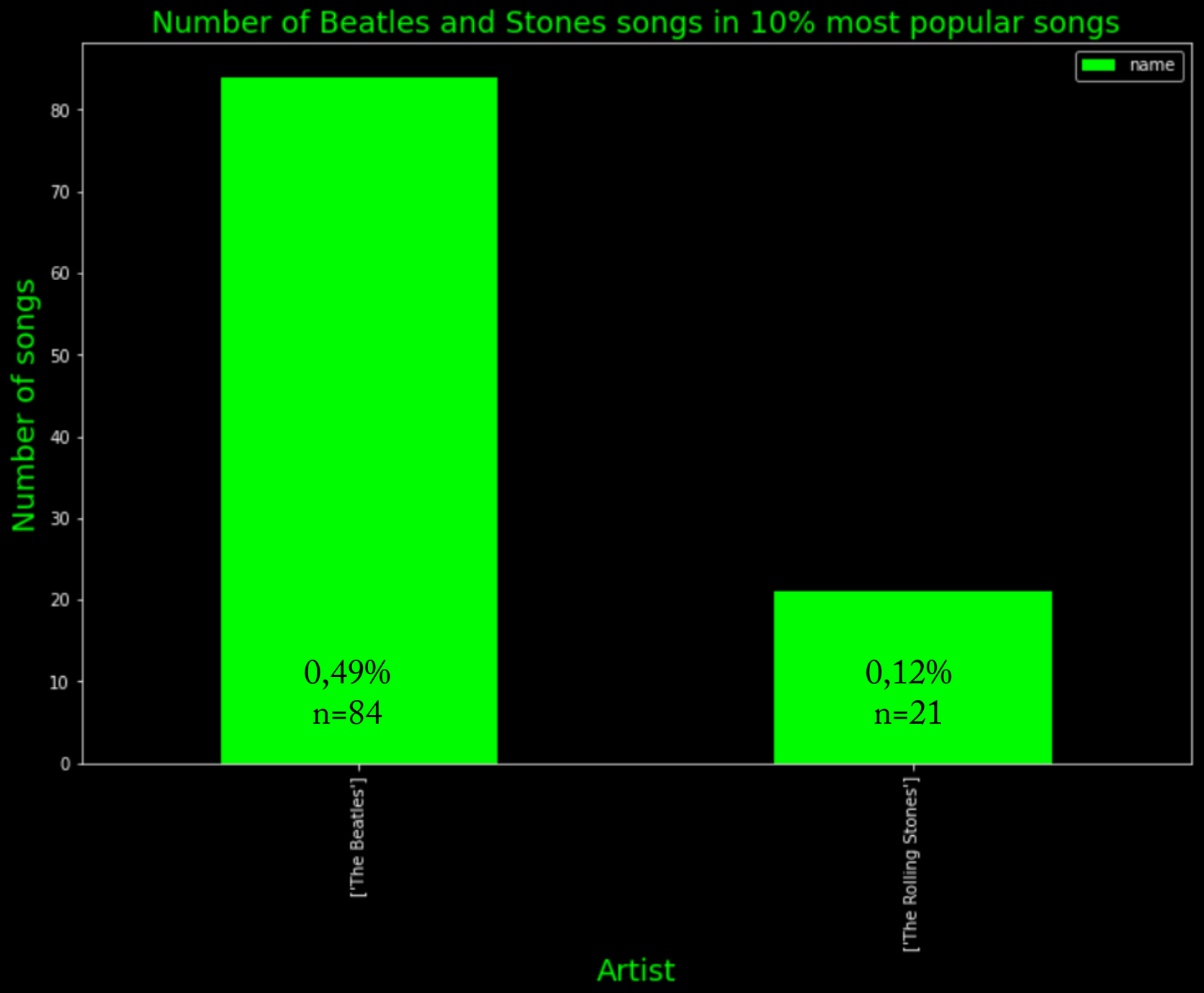


Beatles & Stones averages



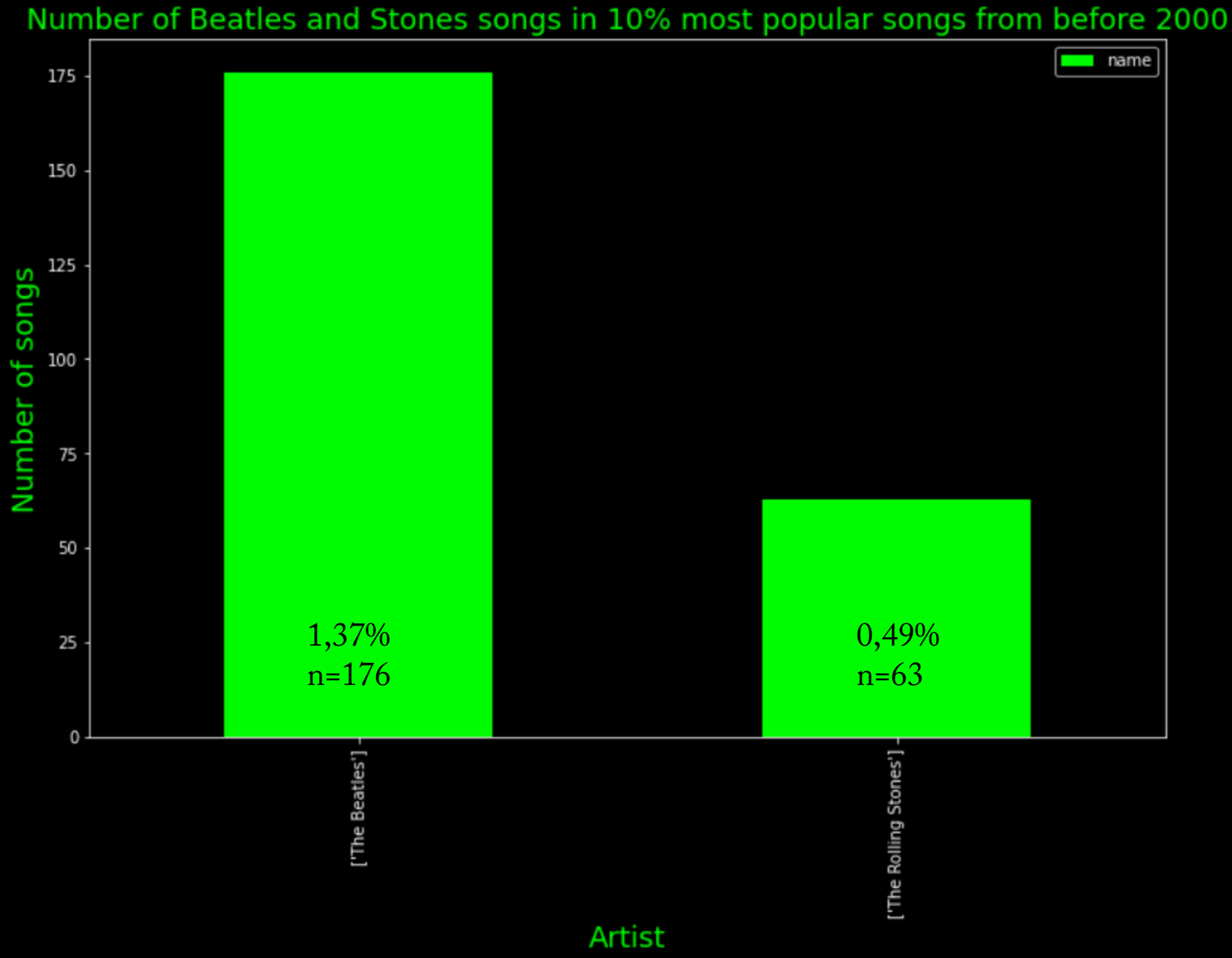
Beatles & Stones Presence in top 10 % popularity

10 % of tracks equals 16991 tracks



Looking at 10 % most popular tracks excluding the 21th century

10 % of tracks equals 12825 tracks



Typical 2019

- I chose the 100 tracks from 2019 that had the highest popularity score
- Then I got their mean on all other parameters
- Then I checked the dataset if there are any other songs that are from before the year of 2000 that sort of match the average 2019 top-100 song
- When finding the matches I had a margin of 0.1 for all 0-1 parameters

Have a listen!

Typical 2018 & 2017

- I found two other datasets with the 100 tracks that had the most listenings in 2018 and 2017
- As with 2019 I wanted to see if there were any matches in the dataset that agreed with the average top-100 track of 2018 or 2017

Have a listen!

We learned that in this dataset:

- Blinding Lights is the most popular song
- BTS and Tones and I & Juice World are the most popular artists
- There's a cyrillic Ernest Hemingway on Spotify
- Explicit songs have increased and acoustic ones decreased over the years
- Danceability and Energy have increased
- The Beatles are generally more appreciated than the Rolling Stones
- And finally that you can't use Spotify's API parameters to find a typical *** song

Thank You!