A descriptive view on Spotify playlists

Linus Stenzel
Matrikelnummer 6008798
linus.stenzel@student.uni-tuebingen.de

Jonas Schmiegel
Matrikelnummer 6002387
jonas.schmiegel@student.uni-tuebingen.de

Abstract

The aim of this paper is to get an intuition of how Spotify playlists are created and used. We take a look at different key data of playlists e.g. tracks, artists and followers. To cover a wide range of topics we chose a descriptive statistics view. Starting with some general stats, we calculate averages of interesting metrics and show how unique values behave when considering more data. Then we analyze playlists with either more mainstream or uncommon tracks. There we found differences in the number of unique artists, but also in the amount of users that give a description of their playlist.[1]

1 Introduction

Music is an essential part of the human culture. Wherever historians found evidence of human settlement, they also found evidence of musical activities. The oldest discovery of an instrument is a bone flute which is determined to be between 43,000 and 82,000 years old.[4, p. 63]

One of the biggest companies in today's music industry is Spotify - a streaming service used by hundreds of millions of people.[2] Since Spotify has the highest market share ($\approx 31\%$) in the music industry [3], we think that our analysis gives an insight of the listening habits of more people than the dataset is representing.

2 The Dataset

The dataset we are using is the *"The Million Playlist Dataset"* [2] from Spotify[3]. It consists of one million playlists and is about 34 GB in size. The dataset was sampled from over four billion public playlists on Spotify. It only includes playlists which were created between January 2010 and November 2017 by US Spotify users, that are at least 13 years old. Originally the dataset was introduced as part of the *RexSys Challenge 2018*. The goal of that challenge was to create an algorithm that suggests appropriate songs to be added to an existing playlist.[4]

[1] The code for this project can be found on https://github.com/JSchmiegel/DataLiteracyProject
[2] In the last shareholder letter of Q4 2021 Spotify announced an active user base of 406 million users.[5]
[3] The dataset can be downloaded from https://research.atspotify.com/datasets/
[4] Further information: https://www.recsyschallenge.com/2018/

3 General Statistics

To get a broad overview of the dataset we decided to make some general numerical analysis on it. The part of the dataset that we are analyzing (=100,000 playlists[5]) contains a total number of 681,805 unique tracks by 110,063 artists in 271,413 albums. The total length of all playlist is 18,124 days 5:59:16 (hrs.:min.:sec.). The average number of tracks in one playlist is ≈ 67[6]. The average duration of a playlist is 4:20:59. This fits with the fact, that a track has an average duration of 0:03:54.[7]

Knowing that the average number of tracks per playlist is 67, we wanted to know how much these contribute to the total amount of unique tracks. For that we looked how the number of unique tracks changes when considering more playlists and plotted the results. To clearly show the relation, we changed the x-axis to a (square) root scale.

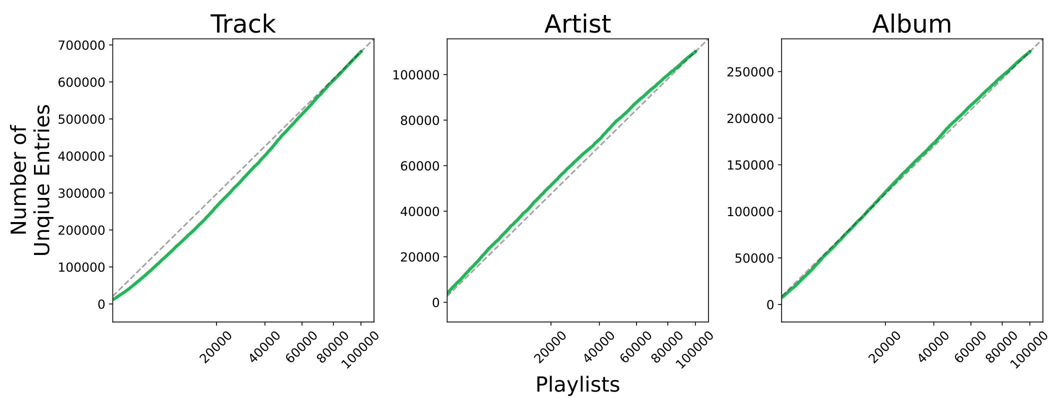

Figure 1: Line plot (—) showing the relation between e.g. unique tracks and considered playlists, with a 45° line (- - -) as baseline

We now know that the number of unique tracks grows asymptotically to $\sqrt{x} \cdot C$, where C is the number of average added unique tracks and x the number of playlists. This also holds for the number of artists and albums.

To see what is mainstream, we got the most popular tracks, albums and artists and counted how often they were occurring in a playlist. The result is shown in tab. 1, where the top 15 tracks, albums and artists are listed according to their popularity. The column "Occur." shows, in how many playlists the corresponding track, artists or album appeared. When comparing the calculated top 15 tracks with the top 15 of the "Year-end Charts Hot 100 Songs 2017" by billboard there is an overlap, of about 40%.[8][1] This is remarkable because the dataset is from January 2010 until November 2017 whereas the billboard top list is only resembling the year 2017.

[5]Originally there are 1 million playlists. Due to limited processing power we had to reduce our analysis to a dataset of 100,000 playlists.

[6]The exact number is 66.778 tracks

[7]Using the following calculation:
average duration of track [ms] $\cdot$ average number of tracks in a playlist $= 234,498.661$ ms $\cdot$ $66.778 = 15,659,351.561$ ms $=$ 4:20:59

[8]The songs "Humble" by Kendrick Lamar, "Closer" by The Chainsmokers, "Bad and Boujee" by Migos, "Congratulations" by Post Malone, "XO TOUR Llif3" by Lil Uzi Vert and "Mask Off" by Future are in both top lists.

Table 1: Top 15 Popular Tracks, Albums and Artists

Pos.	Occur. [%]	Track	Occur. [%]	Album	Occur. [%]	Artist
1	4.44	HUMBLE.	15.38	Stoney	82.43	Drake
2	4.24	One Dance	13.71	DAMN.	40.86	Kanye West
3	4.15	Closer	13.08	Coloring Book	34.55	Kendrick Lamar
4	4.07	Broccoli (feat. Lil Ya...	12.04	American Teen	33.35	Rihanna
5	3.94	Congratulations	11.85	Culture	31.15	The Weeknd
6	3.52	Caroline	11.45	Beauty Behind The Madness	28.44	Eminem
7	3.48	iSpy (feat. Lil Yachty)	11.42	More Life	27.47	Ed Sheeran
8	3.46	XO TOUR Llif3	11.22	The Life Of Pablo	24.97	Future
9	3.44	Location	11.20	Purpose	24.07	J. Cole
10	3.38	Bad and Boujee (feat. ...	11.11	Views	23.62	Justin Bieber
11	3.22	No Role Modelz	10.82	2014 Forest Hills Drive	23.30	Beyoncé
12	3.20	Ignition - Remix	10.46	Blurryface	22.19	The Chainsmokers
13	3.19	Bounce Back	10.13	÷	20.80	Chris Brown
14	3.14	Mask Off	9.90	x	20.59	Luke Bryan
15	3.11	No Problem (feat. Lil ...	9.84	Montevallo	19.85	Calvin Harris

4 Characteristics of Mainstream Playlists

We analyzed the difference between playlists that contain an above-average amount of mainstream tracks and playlists that contain mostly uncommon tracks. For this we first had to establish a measurement from which we can identify a playlist as mainstream. As previously described we have calculated the most common tracks. We determined an arbitrary but reasonable border, which is to contain at least three of the top 50 common tracks to count as a mainstream playlist. To obtain the same group sizes between mainstream and uncommon playlists, we sampled from the larger one (= uncommon playlists). We then looked for differences between the groups and found that playlists with popular tracks have less unique artists than playlists with unpopular tracks. To show this we plotted the number of artists against its occurrence for both groups.

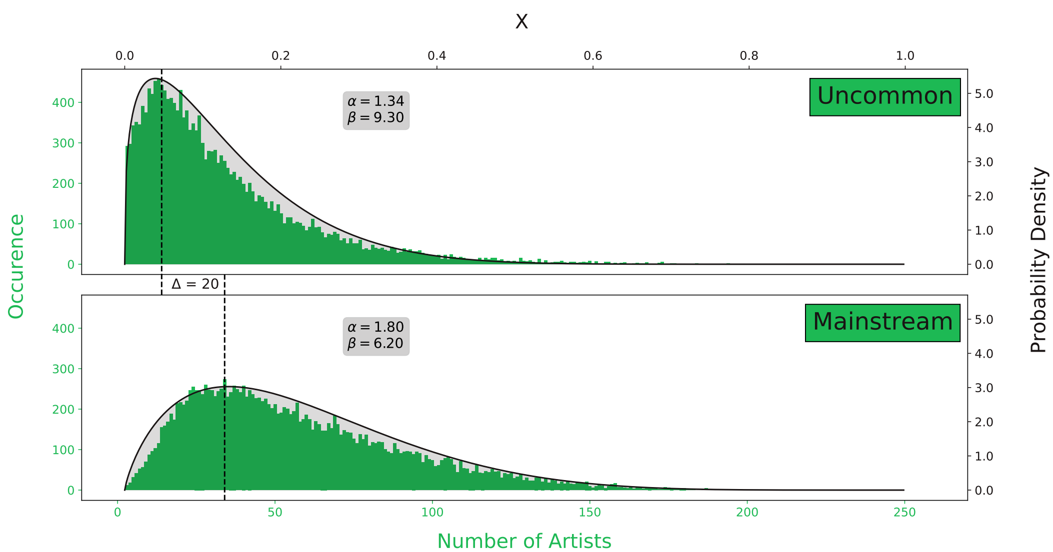

Figure 2: Histogram of Number of Artists (—) with fitted Beta distribution (—) for uncommon and mainstream playlists

Just by looking at the mode (*Uncommon* = 14; *Mainstream* = 34) we can see that there is a difference but to further underlie our finding we fitted a beta distribution to the data. The input values of a beta distribution are between 0 and 1. Since the playlists in the dataset are constrained to have at least three but no more than 250 artists, also the data has limits on both sides. The probability density of both beta distributions is shifted to the lower border. We can see that the beta distribution of uncommon playlists has nearly all of its mass around the mode, while the mass of the beta distribution of mainstream playlists is much more spread.

We also thought of a metric that can directly measure the popularity of a playlist, which is the number of followers of a playlist. The playlists are constrained to have at least one follower and more than 90% of the playlist have three or less. This is why we decided to not split the data into the binary groups, popular and unpopular. Nevertheless, we found that the more follower a playlist has, the more likely it is for the playlist to have a description.

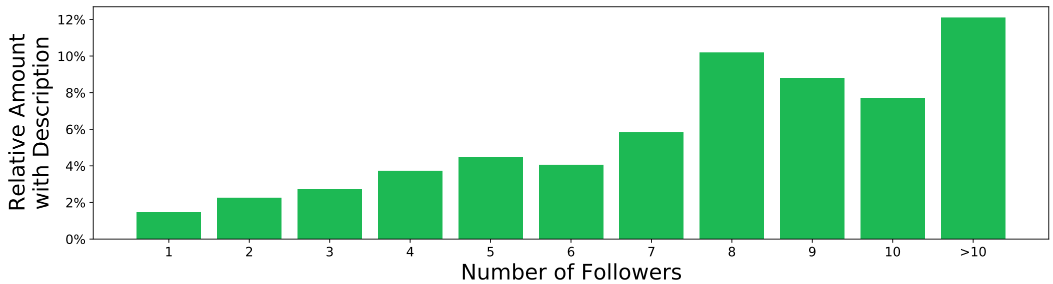

Figure 3: Bar plot of number of followers to relative amount with description

We grouped the playlists by their follower number and counted the ones that have been given a description. Because the groups heavily differ in size, we calculated the relative amount. As mentioned, the data pool for playlists with higher follower numbers gets very small. That is why we put all playlists with more than ten followers into a single group.

5 Conclusion

After analyzing 100,000 playlists we are able to get a feeling how the average playlist would look like. It would have around 67 tracks and lasts about 4:20:59 in total. If the playlist contains at least 3 of the top 50 common tracks (e.g. in year 2017: "HUMBLE", "Closer" and "Congratulations") we can assume that it has around 34 unique artists. Vice versa when counting around 14 unique artists we can assume that this playlists contains less than three common tracks. When we see the playlist has been given a description it is likely that it has more followers than a playlist without a description. Of course, all these measures are no guarantee for the attributes of a playlist, but our paper suggests that the majority of playlists approximately follow these rules.

References

[1] LLC. Billboard Media. Year-end chartshot 100 songs. URL https://www.billboard.com/charts/year-end/2017/hot-100-songs/.

[2] Ching-Wei Chen, Paul Lamere, Markus Schedl, and Hamed Zamani. Recsys challenge 2018: Automatic music playlist continuation. In *Proceedings of the 12th ACM Conference on Recommender Systems*, RecSys '18, page 527–528, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450359016. doi: 10.1145/3240323.3240342. URL https://doi.org/10.1145/3240323.3240342.

[3] Mark Mulligan. Music subscriber market shares q2 2021. URL https://www.midiaresearch.com/blog/music-subscriber-market-shares-q2-2021.

[4] Isabelle Peretz. *The cognitive neuroscience of music*. Oxford University Press. ISBN 0198525192. URL https://www.researchgate.net/profile/Robert-Zatorre/publication/232521640_The_Cognitive_Neuroscience_of_Music/links/02e7e53137ab3a3c8d000000/The-Cognitive-Neuroscience-of-Music.pdf.

[5] Spotify Technology S.A. Shareholder-letter-q4-2021. URL https://s22.q4cdn.com/540910603/files/doc_financials/2021/q4/Shareholder-Letter-Q4-2021_FINAL.pdf.