

Projektbericht: Spotify Analyse

1st Eric Kaufmann
Jena, Germany
eric.kaufmann@uni-jena.de

2nd Maria Gogolev
Jena, Germany
maria.gogolev@uni-jena.de

Abstract—In diesem Projektbericht wurden zwei Spotify-Datensätze von Kaggle verbunden und anschließend analysiert. Dabei wurden Merkmale von Songs durch die Einteilung in Playlisten betrachtet. Mittels Erstellung einer neuen Metrik zur Bestimmung der Populartät können Eigenschaften populärer Songs bestimmt und verglichen werden. Zusätzlich wurden verschiedene Eigenschaften im Laufe der Jahre (1900 - 2020) beurteilt. Somit konnte gezeigt werden, dass Musik energetischer und intensiver wurde und dies große Populartät brachte.

I. EINLEITUNG

Spotify hat sich über die Jahre zu den größten Musik-Streaming-Anbietern der Welt hochgearbeitet und genießt eine sehr große Popularität. Es gibt mittlerweile kaum noch ein Song, welcher nicht auf Spotify zu finden ist. Somit gab es über die letzten Jahre eine Revolution der Musikindustrie. Während man vor einigen Jahren noch Alben der Lieblingskünstler physisch kaufen musste, kann man nun für wenig Geld im Monat Musik aller Artisten aus aller Welt hören. Die selbstgebrannten CDs mit der Lieblingsmusik verschiedener Künstler wurden durch Spotify Playlisten ersetzt, welche man ohne Aufwand teilen und erstellen kann. Ein Merkmal, welches Spotify von seinen Konkurrenten abhebt, ist die sehr gute Empfehlung neuer Songs. Anhand der Songs, welche man gerne hört, findet Spotify ähnliche Musik, sodass man immer mit etwas Neuem konfrontiert wird. Um dies zu erreichen, werden die Eigenschaften der Lieder miteinander verglichen und Ähnlichkeiten hergestellt. Anhand der Eigenschaften kann man herausfinden, was populäre Songs auszeichnet oder was besonders an Künstlern ist. In einer Analyse mittels PySpark¹ versuchen wir herauszufinden, was diese Eigenschaften sind und wie sich diese über die Zeit verändert haben.

II. DATENSÄTZE

Als Datensätze wurden zwei von kaggle.com verwendet:

- 1) Spotify 1.2M+ Songs²
- 2) Spotify Playlists³

Ersteres besteht, wie der Name es schon verdeutlicht, aus ca. 1.2 Millionen (1,204,025 genau) unterschiedlichen Spotify Songs. Zu jedem Song sind dabei 24 Eigenschaften gegeben. Eine Auflistung aller Eigenschaften mit Beschreibung ist in Tabelle (I) zu finden. Der Datensatz liegt als CSV vor und umfasst ungefähr 346MB. Die Daten wurden dabei mittels

der offiziellen Spotify-API generiert. Jeder Song kann mittels einer eindeutigen ID beschrieben werden. Durch eine GET-Anfrage auf <https://api.spotify.com/v1/tracks/{id}> können Informationen zum Album, Songtitel, Artisten und Publikationsdatum extrahiert werden. Mittels zweiter GET-Anfrage auf <https://api.spotify.com/v1/audio-features/{id}> können die restlichen Eigenschaften, wie *danceability*, *energy* oder *liveness* bestimmt werden.

TABLE I
BESCHREIBUNG DER SPALTEN DES SPOTIFY 1.2M+ SONGS DATENSATZ

Eigenschaft	Beschreibung
id	Song-ID
name	Songtitel
album	Albumtitel
album_id	Album-ID
artists	Liste der Artisten
artists_ids	Liste der Artisten-IDs
track_number	Tracknummer des Songs im Album
disc_number	Albumnummer
explicit	Song ist explicit
danceability	Eignung des Songs zum tanzen ($\in [0, 1]$)
energy	Intensität und Aktivität des Songs ($\in [0, 1]$)
key	Tonart
loudness	Lautstärke (dB)
mode	Modus (Dur/Moll)
speechiness	Sprachanteil ($\in [0, 1]$)
acousticness	Akustik eines Songs ($\in [0, 1]$)
instrumentalness	Instrumentalanteil ($\in [0, 1]$)
liveness	Wahrscheinlichkeit einer Liveübertragung ($\in [0, 1]$)
valence	Wertigkeit ($\in [0, 1]$, 0=negativ und 1=positiv)
tempo	Tempo (BPM)
duration_ms	Dauer (ms)
time_signature	Taktart
year	Veröffentlichungsjahr
release_date	Veröffentlichungsdatum (YYYY-MM-DD)

Der Datensatz ermöglicht es Songs anhand von Eigenschaften zu sortieren. Beispielsweise kann man verschiedene Eigenschaften wie die *danceability*, *valence* oder *energy* der Songs über die Zeit betrachten. Eine andere exemplarische Anwendungsmöglichkeit ist ein Algorithmus zur Songempfehlung, bei dem man Songs mit ähnlichen Eigenschaften versucht zu verbinden.

Um die Songs besser unterteilen zu können wurde in dieser Analyse ein zweiter Datensatz verwendet, der Spotify Playlist Datensatz. Dieser besteht aus ca. 2.8 Millionen unterschiedlichen Songs, die in ca. 162 Tausend Playlists von ungefähr 16 Tausend User eingeteilt sind. Trotz einer Größe von ca. 1.8 GB ist der Aufbau des Datensatzes recht einfach. Wie in Tabelle (II) erkennbar besitzt der Datensatz

¹Projekt verfügbar unter https://github.com/eric-kaufmann/spotify_analysis

²siehe <https://www.kaggle.com/datasets/rodolfofigueroa/spotify-12m-songs>

³siehe <https://www.kaggle.com/datasets/andrewmvd/spotify-playlists>

nur vier Spalten. Dabei hat ein User mit einer *user_id* eine Playlist mit der Playlistbezeichnung *playlistname* angelegt. Songs, bestimmt durch den *trackname* und *artistname*, können somit einer Playlist zugehörig sein. Selbstverständlich können dabei Songs mehrfach vorkommen, wenn sie beispielsweise in unterschiedlichen Playlists enthalten sind oder in einer Playlist sogar mehrfach aufkommen.

TABLE II
BESCHREIBUNG DER SPALTEN DES SPOTIFY PLAYLIST DATENSATZ

Eigenschaft	Beschreibung
<i>user_id</i>	User-ID eines Spotify Nutzers
<i>artistname</i>	Name des Artisten
<i>trackname</i>	Songtitel
<i>playlistname</i>	Playlistbezeichnung

Mögliche Anwendungsfelder dieser Playlist ist beispielsweise die Untersuchung der Popularität einiger Songs durch die Anzahl der Aufkommen des Liedes in unterschiedlichen Playlisten. Mehr dazu später.

Um die Songeigenschaften des ersten Datensatzes mit den Playlisteinteilungen des zweiten Datensatzes zu verbinden, müssen die beiden Datensätze verbunden werden. Dafür versuchen wir an den Playlist-Datensatz mittels left-join die Eigenschaften anzuheften. Dabei sollen die gemeinsamen Schlüssel der Songtitel mit Artisten sein. Um dies zu erreichen müssen jedoch vorher die Daten vorbereitet werden. Dafür wurden zunächst die Spaltennamen *name* und *artists*, beim Spotify 1.2M+ Datensatz, bzw. *trackname* und *artistname*, beim Spotify Playlistdatensatz, vereinfacht und in *track* und *artist* umbenannt. Somit sind die beiden Schlüssel gleich bezeichnet.

Der nächste Schritt ist es die Liste von Artisten, welche in jeder Zeile der *artist*-Spalte beim Spotify 1.2M+ Songs Datensatz zu finden ist, aufzusplitten. Diese haben nämlich eine folgende Form:

[*'artist₁'*, *'artist₂'*, ..., *'artist_n'*].

Ziel ist es, dass die Zeile des Datensatzes *n*-Mal repliziert wird und in jeder neuen Zeile nur der jeweilige *artist_i* steht. Zum Schluss wurden die Datentypen aller Spalten angepasst und unrealistische Werte (z.B. Veröffentlichungsjahr 0) aussortiert.

Nach dieser Vorbereitung der Daten ist es nun möglich den left-join des Spotify 1.2M+ Songs Datensatz an den Spotify Playlist Datensatz über die gemeinsame Schlüssel *track* und *artist* durchzuführen. Zur Vereinfachung der späteren Analyse wurde anschließend alle Zeilen mit NA-Werten aus dem verbundenen Datensatz entfernt. Somit werden alle Zeilen entfernt, welche im Spotify Playlist Datensatz keine Verbindung zum Spotify 1.2M+ Songs Datensatz gefunden haben.

Die Verbindung der Datensätze hat jedoch einen großen Datenverlust zur Folge. Nach der Vorverarbeitung der einzelnen Datensätze sind beim Spotify 1.2M+ Songs Datensatz noch ca. 1,194,000 und beim Spotify Playlist Datensatz noch ca. 2,795,000 unterschiedliche Songs übrig. Nach dem join

sind jedoch nur noch ca. 210,000 unterschiedliche Songs übrig. Die Gründe dafür sind recht unterschiedlich. Zum einen gibt es natürlich Songs, die in dem Spotify Playlist Datensatz vorkommen, jedoch nicht im Spotify 1.2M+ Songs Datensatz. Ein anderer Grund sind die Differenzen in der Bezeichnung der Werte der gemeinsame Schlüssel *track* und *artist*. Auf Spotify gibt es sehr häufig Songs, welche zwar einen bestimmten Artisten haben, jedoch ein anderer User dieser Song hochgeladen hat. Somit existiert zwar dieser Song, der Schlüssel *artist* stimmt aber nicht überein. Generell werden zu jedem bekannten Song viele unterschiedliche Versionen von vielen Nutzern hochgeladen. So gibt es von einem Song eine offizielle Version, ein Radio-Edit, eine Live-Version und verschiedene Remix. Auch Features mit anderen Artisten bringen häufig Probleme. Somit ist es möglich, dass ein Song im einen Datensatz die Form "*song_titel*" hat und im anderen "*song_titel (feat. artist2)*". Im Spotify 1.2M+ Songs Datensatz sind die einzelnen Artisten als Liste gespeichert. Beim Spotify Playlist Datensatz sind diese zusammen als String gespeichert. Daraus entsteht das Problem, dass die Artisten im Playlist Datensatz die Form "*artist1, artist2 and artist3*" bzw. "*artist1 & artist2*" haben können. Eine Teilung an Kommata ist aber auch nicht möglich, da es Artisten gibt, welche Kommata in ihrem offiziellen Namen haben. Somit wird die Schnittmenge der beiden Schlüsselmengen noch kleiner. Mit den restlichen Daten können jedoch Einsichten in Eigenschaften von Songs gegeben werden.

III. ANALYSE UND INTERPRETATION

In diesem Kapitel werfen wir einen Blick auf interessante Statistiken der Musik auf Spotify. Speziell werden wir uns anschauen wie sich Musik im Laufe der Jahre verändert, und welche Eigenschaften beliebte Lieder haben.

A. Entwicklung der Musik über die Zeit

Abbildung (1) zeigt wie viele Lieder nach dem Datenverlust pro Jahr übrig bleiben. Wir stellen fest, dass es für einige Jahre keine Einträge mehr gibt, und dass generell viel weniger ältere als moderne Lieder für unsere Analyse zur Verfügung stehen. Für die folgende Analyse ist es also wichtig zu erwähnen, dass die Ergebnisse für ferner in der Vergangenheit liegende Musik weniger zuverlässig sind. Aus den vorliegenden Daten geht hervor, dass sich die Musik im Laufe der Jahre in Bezug auf verschiedene Attribute deutlich verändert hat, und gewisse Trends ersichtlich sind. Die Analyse bezieht sich auf Abbildung (2). Im Folgenden werden nicht alle Attribute aufgegriffen, da bei ihnen Statistiken wie Standardabweichung und Durchschnittswert weniger aussagekräftig sind (beispielsweise *key* also der Notenschlüssel).

- *danceability* Beginnend mit der *danceability* sehen wir, dass der Minimal- und Maximalwert über die Jahre leicht auseinander gehen. Nach einem Absinken der *danceability* in den 1940 Jahren zeigt der Durchschnittswert wieder einen leichten Aufwärtstrend. Die Standardabweichung bleibt relativ stabil über die Zeit.

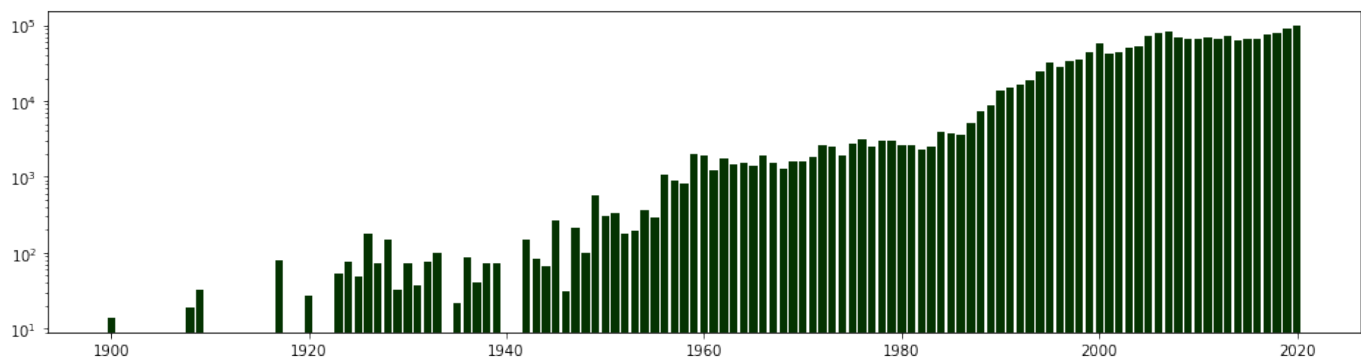


Fig. 1. Eine logarithmisch skalierte Darstellung der Anzahl der Lieder pro Jahr.

- *energy* Auch bei der Energie ist ein ähnlicher Trend zu beobachten: Der Minimalwert bleibt relativ konstant, der Maximalwert steigt stetig an, der Durchschnittswert zeigt einen Aufwärtstrend und die Standardabweichung nimmt im Laufe der Zeit zu. Dies deutet darauf hin, dass die Musik im Laufe der Jahre energiegeladener geworden ist, es gleichzeitig aber mehr Diversität in diesem Attribut gibt.
- *loudness* Bei der Lautstärke ist ein sinkender Minimalwert und eine leichte Zunahme des Maximalwerts und Durchschnittswertes zu verzeichnen. Musik wird also leicht lauter.
- *mode* Der Durchschnittswert des Modus bleibt über die gesamte Zeit überhalb von 0.5, was aussagt, dass stets mehr Musik in Dur produziert wird als in Moll, wobei ein leichter Abstieg des Durchschnittswertes in den letzten Jahren zu verzeichnen ist. Es wurde also in den letzten Jahren mehr in Moll produziert.
- *speechiness* Das Minimum des Sprachanteils bleibt nahe bei 0, da Instrumentalmusik früher wie auch heute relevant ist. Der anfangs geringe Maximalwert entwickelt sich hingegen nach oben. Durchschnittswert und Standardabweichung bleiben gering was darauf hindeutet dass der Großteil der Musik nach wie vor instrumental ist. Ansonsten zeigt die *speechiness* keinen starken Trend.
- *acousticness* Die Akustik zeigt über die Jahre einen Rückgang der Durchschnittswerte, was darauf hindeutet, dass die Musik im Laufe der Jahre elektronischer geworden ist. Die Standardabweichung der *acousticness* nimmt im Laufe der Zeit zu, was darauf hindeutet, dass die Verteilung der Akustizitätswerte im Laufe der Zeit immer vielfältiger und unterschiedlicher wird. Elektronische Musik ersetzt akustische Musik also nicht, sondern kommt zusätzlich hinzu.
- *instrumentalness* Bei der Instrumentalität ist ein Auf-und-Ab zu sehen, ohne klaren Trend hinsichtlich wachsendem oder fallendem Instrumentalanteil. Die Standardabweichung bleibt konsistent hoch, was bedeutet, dass nach wie vor viele Lieder mit hohem als auch niedrigem Instrumentalanteil produziert werden.
- *liveness* Auch *liveness* zeigt keinen eindeutigen Trend. Der Durchschnittswert ist relativ niedrig mit einer geringen Standardabweichung, was darauf hindeutet dass der Großteil der Musik nicht live übertragen wurde.
- *valence* Die Valenz zeigt keinen konstanten Abstieg der Statistiken, aber einen Reduktion des Durchschnittswerts ab den 1940ern ähnlich zur *danceability*, was darauf hindeutet, dass die Musik seit dieser Zeit weniger positiv ist.
- *tempo* Das Maximaltempo wird über die Jahre höher und das Minimaltempo wird niedriger während der Durchschnittswert relativ konstant im mittleren Bereich der Attributskala bleibt. Die Standardabweichung ist niedrig und konstant. Die meisten Lieder sind also weder langsam noch schnell, aber es wird über die Jahre immer wieder mit neuen Tempo-Rekorden experimentiert.
- *duration* Die Dauer zeigt einen starken Anstieg des Maximalwertes. Minimal- und Durchschnittswerte bleiben relativ konsistent und gering, genauso wie die Standardabweichung. Die Musik ist in Bezug auf die Dauer eher stabil und vorhersehbar.
- *time signature* Der Anteil der Songs mit $\frac{4}{4}$ -Takt ist im Laufe der Zeit relativ stabil und stark geblieben. Die Standardabweichung der Taktart ist klein und bleibt im Laufe der Zeit stabil, was darauf hindeutet, dass die meisten Songs hauptsächlich 4 Schläge pro Takt haben.

Insgesamt erweckt die Grafik den Anschein, dass die Musik im Laufe der Jahre energiegeladener und etwas tanzbarer und lauter geworden ist, wobei die durchschnittliche Tanzbarkeit und das Energieniveau allmählich gestiegen sind. Die durchschnittliche Akustik und Instrumentalität schwanken. Die Musik scheint aber insgesamt zu mehr elektronischen oder synthetischen Klängen zu tendieren, da sich der Anteil der akustischen Musik über die Jahre eindeutig verringert. Für die Attribute deren Durchschnittswert einem klaren Trend folgt, also der relative Rückgang akustischer Elemente und der Anstieg der Energie, lässt sich auch ein Anstieg in der Standardabweichung, also der Diversität in der Ausprägung des Attributs feststellen.

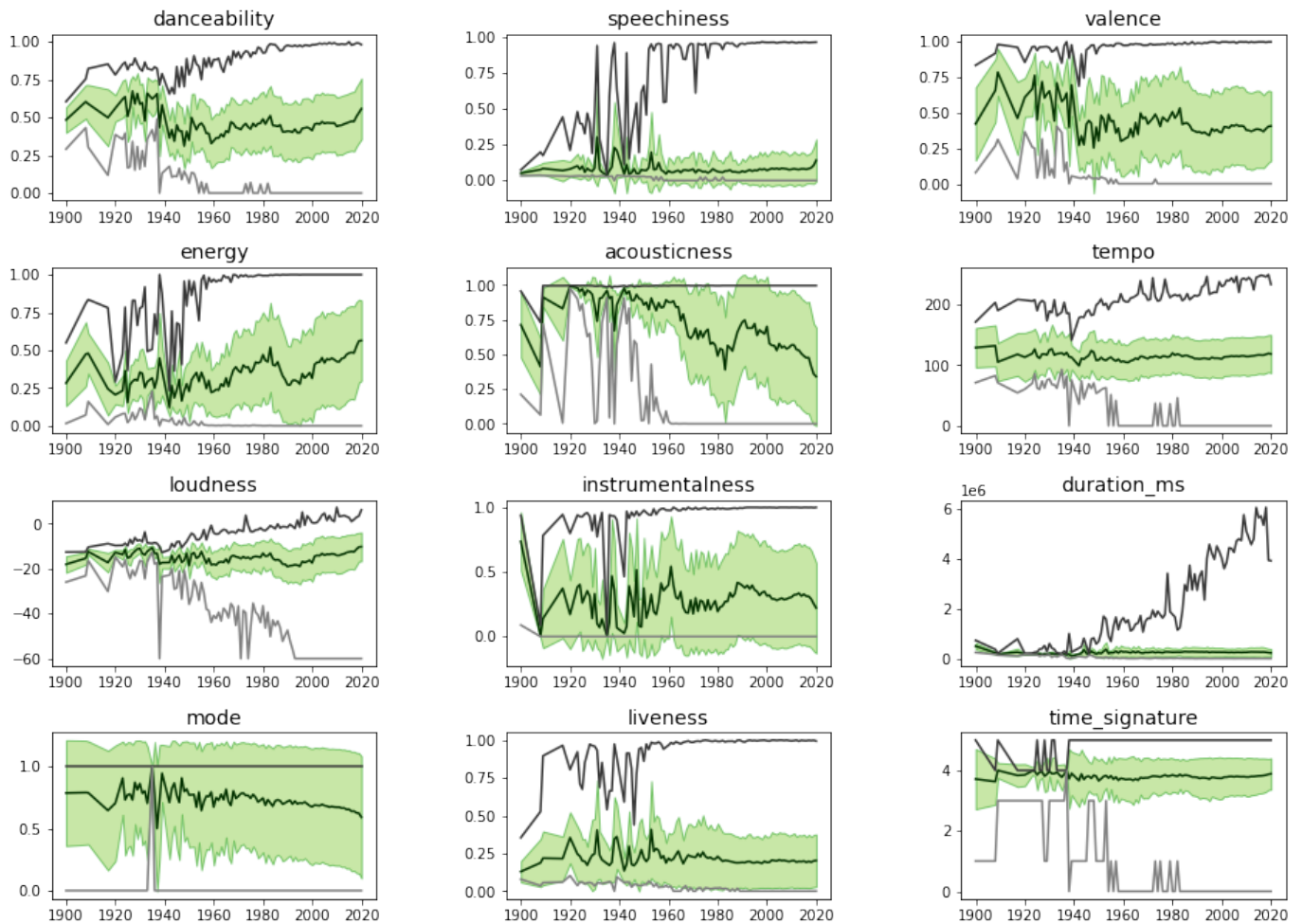


Fig. 2. Entwicklung der Musik-Attribute von 1900 bis 2020. Die in der Mitte verlaufende Linie ist der Durchschnittswert, die obere Linie das Maximum, die untere Linie das Minimum und der grüne Bereich ist die Standardabweichung im gegebenen Jahr.

B. Popularität

Das Ziel in diesem Abschnitt besteht darin, die Statistiken der Attribute durchschnittlicher Lieder und Künstler mit den Statistiken populärer Lieder Künstler zu vergleichen, um herauszufinden was Popularität ausmacht. In keinem der beiden Datensätzen gibt es Informationen über die Beliebtheit, weshalb wir eine neue Metrik dafür einführen. Diese soll sowohl für jeden Künstler als auch für jeden Song gelten. Dabei wird gezählt wie oft ein gegebenes Lied beziehungsweise ein gegebener Künstler in unterschiedlichen Playlists vorkommt. Mehrfache Aufkommen innerhalb einer Playlist werden nicht mitgerechnet. Das ergibt sich einen Schätzwert für die Popularität. Tabelle (III) zeigt die 10 populärsten Lieder und Tabelle (IV) 10 populärsten Künstler auf. Abbildung (3) zeigt wie oft im Durchschnitt ein Lied, welches im gegebenen Jahr veröffentlicht wurde, in Playlists erschien. Es lässt sich feststellen, dass besonders Songs aus den letzten Jahren eine hohe Popularität genießen.

1) *Vergleich populärer Lieder mit durchschnittlichen Liedern:* In Abbildung (4) wird deutlich, dass die Durch-

Songtitel	Artist	Anzahl
Ho Hey	The Lumineers	1547
Kids	MGMT	1319
Rather Be (feat. Jess Glynne)	Clean Bandit	1279
Do I Wanna Know?	Arctic Monkeys	1204
Chandelier	Sia	1169
Don't Stop Believin'	Journey	1114
Sail	AWOLNATION	1110
Take Me Out	Franz Ferdinand	1084
Creep	Radiohead	1066
Summer	Calvin Harris	1058

TABLE III
TOP 100 DER POPULÄRSTEN LIEDER

schnittswerte der populären Songs wesentlich niedriger für *danceability*, für die Lautstärke, für akustische Elemente, für Tempo und für die Taktart. Das heißt populäre Lieder eignen sich im Durchschnitt weniger zum Tanzen, sind viel leiser, langsamer, und haben einen geringeren akustischen Anteil. Sie haben weniger Schläge pro Takt. Populäre Songs sind jedoch energetischer, was logisch erscheint, da die durchschnittliche Energie (siehe Abb. (2)) und die durchschnittliche Beliebtheit

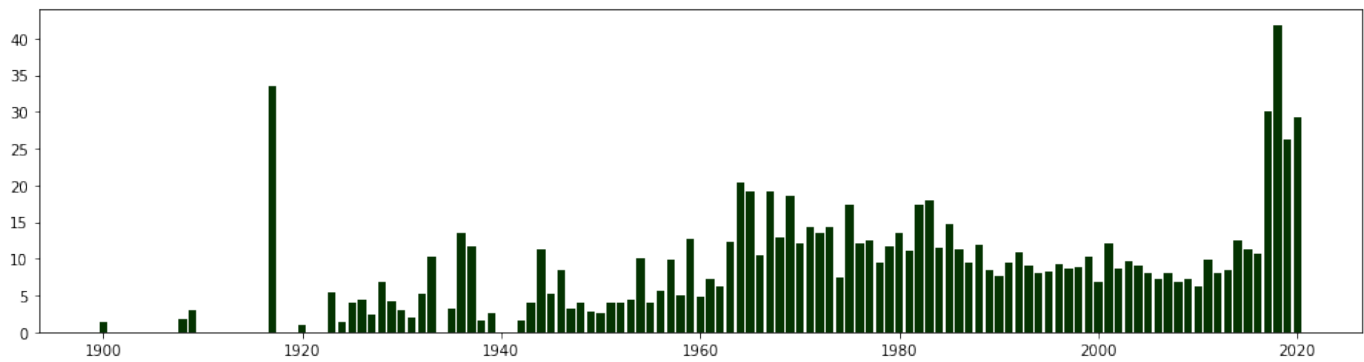


Fig. 3. Durchschnittliche Beliebtheit der im gegebenen Jahr veröffentlichten Musik.

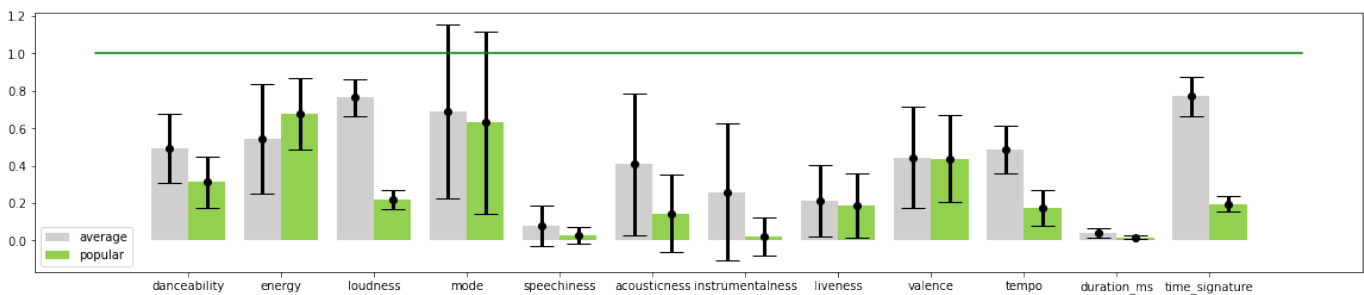


Fig. 4. Vergleich des Durchschnittswerts und der Standardabweichungen verschiedener Attribute der Top 100 der beliebtesten Songs mit allen Songs. Rechts (grün) sind die populären Lieder dargestellt, und links (grau) sieht man die durchschnittlichen Lieder. Da die Attribute unterschiedliche Wertebereiche haben, werden die Durchschnittswerte und die Standardabweichungen jeweils mit dem globalen Maximum und dem globalen Minimum des Attributs skaliert.

Artist	Anzahl
Beyoncé	3476
Radiohead	3406
Arctic Monkeys	3038
Coldplay	2855
Michael Jackson	2593
MGMT	2516
Bob Dylan	2307
Foo Fighters	2212
The Strokes	2189
Bruce Springsteen	2131

TABLE IV
TOP 100 DER POPULÄRSTEN KÜNSTLER

(siehe Abb. (3)) in den letzten Jahren besonders hoch ist. Außerdem kann man erkennen, dass die Standardabweichung für die top 100 beliebtesten Lieder in fast allen Attributen geringer ist, was darauf hindeutet, dass sich populäre Lieder untereinander in ihren Attributen ähnlicher sind. Man kann auf Grund dieser Resultate populäre Lieder (Lieder die oft in Playlisten auf Spotify aufgenommen werden) wie folgt charakterisieren: sie haben trotz geringerem Tempo und geringer Lautstärke viel Energie leicht positiver, und generell weniger Variation in den Attributen. Eventuell eignen sie sich also eher zum Hören im Hintergrund.

2) *Vergleich populärer Künstler mit durchschnittlichen Künstlern:* Zunächst haben wir die Durchschnittswerte für jeden Künstler und jedes Attribut durch Aggregation aller

Lieder des Künstlers über das arithmetische Mittel errechnet. Dannach wurden die top 100 beliebtesten Künstler gefiltert, und mit allen Künstlern insgesamt verglichen.

In Abbildung (5) sieht man viele Ähnlichkeiten zu den Ergebnissen in Abbildung (4).

Ein paar kleinere Unterschiede sind die im Durchschnitt geringere "Valence" der populären Künstler im Vergleich zum Durchschnittskünstler und die Ähnlichkeit der durchschnittlichen Energielevel. Die kleine Standardabweichung ist auch hier wieder indikativ für die Ähnlichkeit der populären Künstler untereinander. Was man in beiden Abbildungen erkennen kann, ist dass Popularität mit weniger Eignung zum Tanzen, langsamerem Tempo, weniger Schlägen pro Takt, geringerer Lautstärke, weniger Instrumentalanteil, weniger akustischen Elementen und weniger Varianz in den Attributen einhergeht.

3) *Vergleich der Eigenschaften populärer Lieder mit den Liedern populärer Künstler:* In Abbildung (6) sieht man die Statistiken beliebter Künstler gegenübergestellt zu den Statistiken beliebter Lieder. Man kann folgende Unterschiede feststellen: Die Varianz der Attribute scheint bei beliebten Künstlern noch stärker eingeschränkt zu sein als bei den beliebtesten Songs, was bedeutet dass sich beliebte Künstler untereinander mehr ähneln als beliebte Lieder. Weiterhin haben beliebte Künstler etwas weniger energetische Lieder, mit etwas höherem Akustik-Anteil.

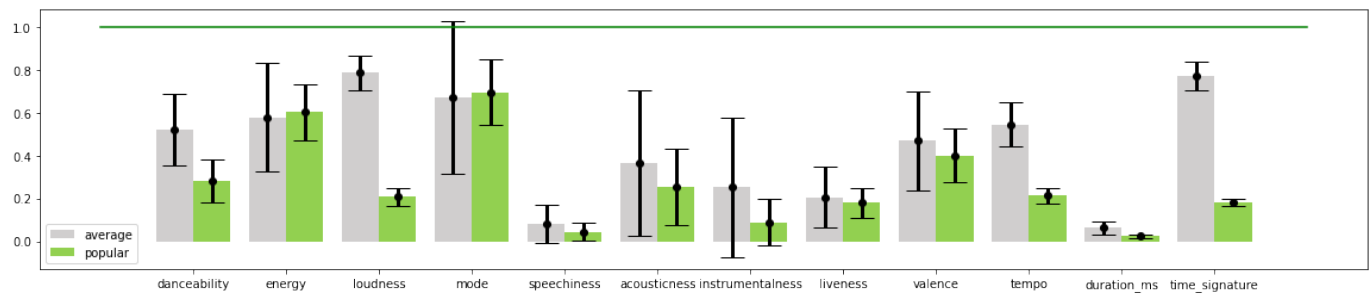


Fig. 5. Vergleich des Durchschnittswerts und der Standardabweichungen verschiedener Attribute der den Top 100 beliebtesten Künstler mit allen Künstlern insgesamt. Rechts (grün) sind die populären Künstler dargestellt, und links (grau) sieht man die durchschnittlichen Artisten. Da die Attribute unterschiedliche Wertebereiche haben, werden die Durchschnittswerte und die Standardabweichungen jeweils mit dem globalen Maximum und dem globalen Minimum des Attributs skaliert.

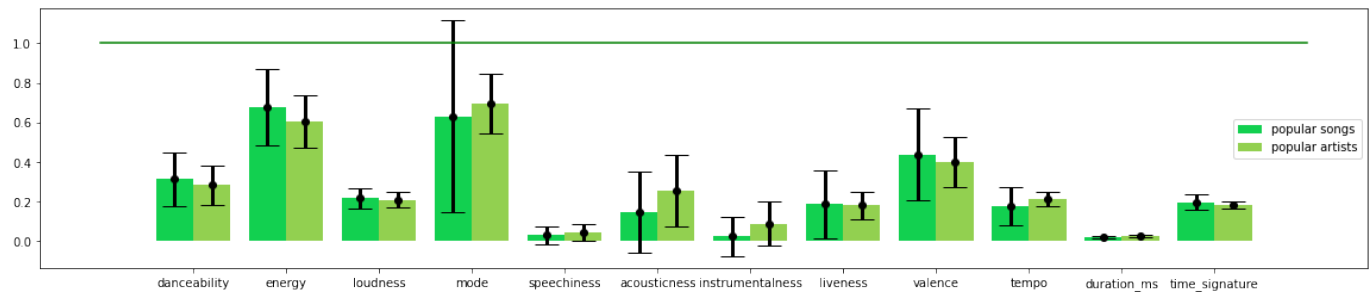


Fig. 6. Vergleich des Durchschnittswerts und der Standardabweichungen verschiedener Attribute der Durchschnittssongs der Top 100 beliebtesten Künstler mit den top 100 beliebtesten Songs insgesamt. Rechts sind die populären Künstler dargestellt, und links sieht man die populären Songs. Da die Attribute unterschiedliche Wertebereiche haben, werden die Durchschnittswerte und die Standardabweichungen jeweils mit dem globalen Maximum und dem globalen Minimum des Attributs skaliert.

C. Zusammenfassung

Die Analyse unterschiedlicher Statistiken zeigte einige interessante Trends und Entwicklungen der Musik über den Zeitraum von 1900 bis 2020. Musik scheint im Laufe der Zeit im Durchschnitt energetischer zu werden, und es gibt einen Trend zu mehr elektronischer Musik. Die Standardabweichung dieser Attribute zeigt aber auch, dass nach wie vor viel akustische Musik und Musik mit niedrigerer Intensität produziert wird. Man sieht auch, dass die Extremwerte hinsichtlich der Lautstärke, des Tempos, und der Dauer immer weiter auseinander entwickeln, also extremer werden. Das alles erweckt den Anschein, dass moderne Musik eine höhere Diversität hat. Das untersuchen der Beliebtheit zeigte hingegen, dass sich sowohl beliebte Songs als auch beliebte Künstler in ihren Attributen weniger unterscheiden. Weiterhin ließ sich feststellen, dass beliebte Musik energetischer und elektronischer ist. Dies befindet sich im Einklang mit der ersten Beobachtung, da wir weiterhin herausgefunden haben, dass die Musik der letzten Jahre besonders populär ist. Außerdem hat populäre Musik weniger Schläge pro Takt, ein geringeres Tempo, einen geringeren Instrumentalanteil, ist leiser, und eignet sich weniger zum Tanzen, wird also wahrscheinlich eher zum Hören im Hintergrund oder zum aktiven Genießen der Musik genutzt.

REFERENCES

- [1] Pichl, Martin; Zangerle, Eva; Specht, Günther: "Towards a Context-Aware Music Recommendation Approach: What is Hidden in the

- Playlist Name?" in 15th IEEE International Conference on Data Mining Workshops (ICDM 2015), pp. 1360-1365, IEEE, Atlantic City, 2015.
- [2] Figueroa, Rodolfo. (2020). Spotify 12M Songs [Dataset]. Kaggle. <https://www.kaggle.com/datasets/rodolfofigueroa/spotify-12m-songs>