

Projekt C) Spotify

Betriebswirtschaftliche Datenanalyse (282135)

Avram Ivanovic (Matr. 209332)

David Pehar (Matr. 207557)

Suphi Pembe (Matr. 207617)

3. Semester

Vorgelegt bei

Dr. Benjamin M. Abdel-Karim

13. Januar 2022

Inhaltsverzeichnis

Abbildungsverzeichnis	3
Abstract	4
1 Einleitung	5
1.1 Motivation	5
1.2 Forschungsfrage	5
1.3 Aufbau der Arbeit	6
2 Theoretische Grundlagen	7
3 Forschungsdesign	10
3.1 Aufbau Analyse und Methodischer Ansatz	10
3.2 Datenbereinigung und Datenbehandlung	13
4 Ergebnisse	18
5 Limitation	23
6 Fazit	25
7 Literaturverzeichnis	27
8 Ehrenwörtliche Erklärung	28

Abbildungsverzeichnis

Abbildung 1	eigene Grafik
Abbildung 2	eigene Grafik
Abbildung 3	eigene Grafik
Abbildung 4	eigene Grafik
Abbildung 5	eigene Grafik
Abbildung 6	eigene Grafik
Abbildung 7	eigene Grafik
Abbildung 8	eigene Grafik
Abbildung 9	eigene Grafik
Abbildung 10	eigene Grafik
Abbildung 11	eigene Grafik
Abbildung 12	eigene Grafik
Abbildung 13	eigene Grafik

Abstract

In dieser Seminararbeit werden zwei Datensätze mit den Daten zu den weltweiten Spotify-Charts der Jahre 2017 bis 2020 mit Python, NumPy, Pandas und Matplotlib bereinigt und anschließend analysiert. Der Ablauf dieser Arbeit und die genutzten Methoden werden dabei mit Diagrammen und Coding-Beispielen aus der genutzten Entwicklungsumgebung „Pycharm“ erklärt. Außerdem werden Daten aus den Datensätzen analysiert, erklärt und interpretiert, um eine Prognose zum Verlauf der Charts von Spotify aufzustellen. Anschließend werden die Limitationen aufgezeigt, um für einen Rahmen zu sorgen, in dem diese Arbeit ihre Aussagekraft besitzt. Zuletzt wird die eigentliche Prognose aufgestellt.

1 Einleitung

Spotify gehört zu den meistgenutzten Streamingdiensten für Musik weltweit. Aufgrund von hoher Nachfrage und einem benutzerfreundlichen Konzept konnte sich Spotify als eine der wichtigsten Plattformen für Musikliebhaber und vor allem Künstler aller Art durchsetzen.¹

1.1 Motivation

Unter den drei Autoren dieser Arbeit befindet sich niemand, der nicht täglich privat von Spotify Gebrauch macht. Zusätzlich ist einer der Autoren Teil von zwei aktiv verfolgten Musikprojekten, die bereits ihren Weg zu Spotify gefunden haben. Allerdings mit milderem Erfolg auf der Plattform als viele andere Künstler. Aufgrund dieses Verhalts besteht ein persönliches Interesse an Spotify und Daten, die Künstler dort hinterlassen.

1.2 Forschungsfrage

Mit den uns bereitgestellten Datensätzen ist es unser Ziel Prognosen für künftig aufsteigende Musiktrends zu treffen. Analysiert werden die Daten der globalen Charts der Jahre 2017 bis 2020. Die beliebtesten Künstler werden miteinander verglichen, um daraus Erfolgsfaktoren zu bestimmen. Allgemein lautet also die Frage, deren Antwort gesucht wird:

„Welche Musiktrends sind in Betracht der Erfolgsfaktoren der beliebtesten Künstler der Jahre 2017 bis 2020 auf Spotify zu erwarten?“

¹Vgl. Gomes, I. (2021) S.337-338

1.3 Aufbau der Arbeit

Diese Arbeit besteht aus weiteren fünf Kapiteln, in denen jeweils verschiedene Thematiken behandelt werden, um die gewünschte Prognose treffen zu können. Im nächsten Kapitel namens: „Theoretische Grundlagen“, sollen die Grundbedingungen dieser Datenanalyse aufgeführt werden. Das bedeutet: im Folgenden werden Begriffserklärungen, Definitionen, und weitere für den Kontext dieser Arbeit relevante Informationen erläutert werden.

Darauf folgt das Kapitel „Forschungsdesign“. Eine Ausführung unserer gesamten Arbeit mit den gestellten Datensätzen. Hier wird der Aufbau unserer Datenbereinigung und Datenanalyse behandelt, und das praktisch Angewendete in der Theorie erklärt. Die Ergebnisse jener Datenanalyse befinden sich im anschließenden gleichnamigen Teil „Ergebnisse“, wo auch der Bezug zum ursprünglichen Ziel aufgebaut, um die erhoffte Prognose aufzustellen.

Das anschließende Kapitel „Limitation“, grenzt diese Arbeit mithilfe der Einordnung der Grundbedingungen in einen allgemeineren Kontext ein. Zuletzt folgt eine Zusammenfassung mit dem Fazit, wodurch sich die Forschungsfrage und somit die Prognose erschließt. Alle Referenzen werden im Weiteren im Literaturverzeichnis aufgelistet.

2 Theoretische Grundlagen

Spotify entstand 2006 in der schwedischen Hauptstadt Stockholm. Gegründet von Daniel Ek und Martin Lorentzon, mit dem Ziel eine Lösung für das Problem der digitalen Raubkopie aus jener Zeit zu finden. Die Lösung stellte sich als das Erschaffen eines Streamingdienstes für Musik heraus.² Streaming bedeutet dabei allgemein ein temporäres Speichern von Daten, sodass Medien in Echtzeit während des Herunterladens konsumiert und anschließend automatisch gelöscht werden. Dabei existieren zwei Tarife: Ein kostenloser und ein monatlich kostenpflichtiger. Ohne ein Abonnement, benötigen Nutzer einen eigenen Account und eine stetige Internetverbindung. Limitationen existieren durch beispielsweise Werbung, niedrige Tonqualität und Einschränkungen bei der Musikkwahl. Mit einem Abonnement fallen eben genannte Punkte weg, und es kann zusätzlich Musik gezielt heruntergeladen werden, um diese auch ohne Internetverbindung zu hören. Spotify existiert als Online-Anwendung und als eigene App für mobile und immobile Geräte der meisten Betriebssysteme. Künstler können mit der App „Spotify for Artists“ zusätzlich viele der von Spotify erhobenen Daten zum eigenen Account einsehen. Da Spotify alle individuellen Daten der Nutzer sammelt, um diese für ein optimales Nutzererlebnis zu nutzen, existiert ein entsprechend großer Datenfluss. Zu diesen Daten gehören z.B. die allgemeine Häufigkeit des Streamens aller Songs, monatlich erreichte Hörer, aber auch eine Auflistung nach Stream-Anzahl pro Zeit und Region, d.h. Beliebtheit und Charts. Diese Daten werden in dieser Arbeit zum Teil behandelt und analysiert, um eine Prognose für mögliche kommende Musiktrends aufzustellen.

Jene Datensätze wurden uns für diese Arbeit über die Plattform „keggel.com“ zur Verfügung gestellt. Der Begriff Datensatz bedeutet hierbei: Eine Gruppe von inhaltlich zusammenhängenden Informationen beziehungsweise Daten wie z.B. Name und Nummer einer Datei. Datensätze entsprechen einer logischen Struktur, die bei der Softwareentwicklung festgelegt

²Vgl. Gomes, I. (2021) S.341-342

wurde. In unserem Fall sind die Datensätze als sogenannte „CSV“-Dateien gespeichert. Diese Abkürzung steht für „comma-separated-values“. CSV-Dateien trennen eine Datenmenge aus Informationen mit Zeichen wie Kommas, Semikolons oder Doppelpunkten.³ Dies eignet sich um große Mengen von Informationen tabellarisch sortiert zu speichern. Mit entsprechender Software, wie beispielsweise Microsoft Excel, können CSV-Dateien bis zu einer bestimmten Größe direkt als Tabelle geöffnet, eingesehen und bearbeitet werden. In unserem Fall handelt es sich um zwei CSV-Dateien, die aufgrund ihrer Größe von etwa zwei Gigabyte lokal im Ordnerpfad unserer Entwicklungsumgebung gespeichert sind.

Unser erster Datensatz, "Database to calculate popularity.csv", beinhaltet folgende Daten:

- Interpreten der Songs
- Songtitel
- Datum der Chartplatzierung der Songs
- Länder Chartplatzierungen der Songs
- Positionen der Songs in den Charts
- URI (einzigartiger Primärkey, d.h. eine Information für eindeutige Zuordnung wie z.B. eine ID-Nummer)

Unser zweiter Datensatz, "Final Database.csv", beinhaltet folgende Daten:

- URI
- Beliebtheit eines Songs als Parameter
- Songtitel
- Interpreten
- Album oder Single-Release als Boolean
- Albumname, im Fall eines Album-Release
- Musikgenre des Songs
- Anzahl der Follower des Künstlers auf Spotify

³Vgl. Dörn, S. (2020) S.83

- Explizität, ob der Song Vulgäre Inhalte enthält als Boolean
- Länge der Songs in Millisekunden

Für die Bereinigung, Bearbeitung und Analyse unserer Datensätze verwendeten wir die Programmiersprache Python in der Version 3.9 verwendet, in der Entwicklungsumgebung Pycharm. Dabei sind zusätzlich die Programmbibliotheken, zu Englisch „Libraries“, „NumPy“, „Pandas“ und „Matplotlib“ zum Einsatz gekommen. NumPy dient hierbei der Handhabung von Vektoren, Matrizen oder anderen Arrays, Ansammlungen von Daten. Außerdem dient NumPy mit seinen Funktionen der numerischen Berechnung. Pandas bietet Funktionen für die Verwaltung und Analyse von Daten, und Matplotlib erlaubt die mathematische Darstellung von Grafiken.

Des Weiteren existieren noch Fachbegriffe, die vor dem praktischen Teil einer kurzen Erklärung bedürfen. Ein „Dataframe“ erfasst Variablen eines Datensatzes als Spalten, und die Beobachtungen als Zeilen. Die Variablen als Spalten eines Datensatzes sind dabei die Parameter bzw. Größen der obersten Zeile der Tabelle. Hier werden die Kategorien den Spalten der Tabelle zugewiesen. Beobachtungen eines Datensatzes sind die einzelnen Zeilen der Tabelle, auch „Rows“ genannt.

3 Forschungsdesign

Im folgenden Abschnitt wird nun ausführlich die Forschung und Analyse behandelt, die den Rahmen dieser Arbeit bildet. Vorab gilt es jedoch für ein allgemeineres Verständnis den Ablauf unserer Forschung aufzuführen.

3.1 Aufbau Analyse und Methodischer Ansatz

Zuerst ist eine konkrete Motivation, um sich mit dem gewählten Thema zu befassen nötig. Unsere Motivation, wie zuvor beschrieben, war dabei größtenteils ein persönliches Interesse, aber auch das Lernen von Fähigkeiten für erfolgreiches Durchführen von Datenanalysen.

Der nächste wichtige Schritt war das Wählen einer zur Thematik und der Motivation passenden Forschungsfrage. Eine Prognose zu möglichen Musiktrends anhand von Daten zu Künstlern aus der Vergangenheit aufzustellen, schien dabei ein ansprechendes Ziel zu sein.

Nach der Wahl der Forschungsfrage war die Recherche für ein ausreichendes Hintergrundwissen ein weiterer Schritt. Dabei nutzten wir die offizielle Python-Dokumentation und die Dokumentation jeder einzelnen genutzten Programmbibliothek. Außerdem informierten wir uns mit den öffentlichen Lernvideos von Dr. Benjamin M. Abdel-Karim zu Methoden der Datenanalyse. Des Weiteren nutzten wir Fachliteratur um Fakten zu finden, und Begriffe zu definieren.

Mit Hintergrundwissen und Zusammenfassungen allein kann ein Vorhaben wie diese Arbeit nicht umgesetzt werden, daher richteten wir eine förderliche Umgebung ein. Pycharm inklusive aller Libraries wurde installiert, eingerichtet, und an ein privates Projekt aller Mitglieder in einem Versionskontrollsystem, wie z.B. GitHub verknüpft.

Nun galt es, sich einen Überblick über den Datensatz durch das Lesen und Überfliegen

der beiden CSV-Dateien zu verschaffen. Dabei hielten wir fest, welche Informationen wir gebrauchen konnten, um eine mögliche Prognose über Musikrends anhand von deren Interpretation aufzustellen.

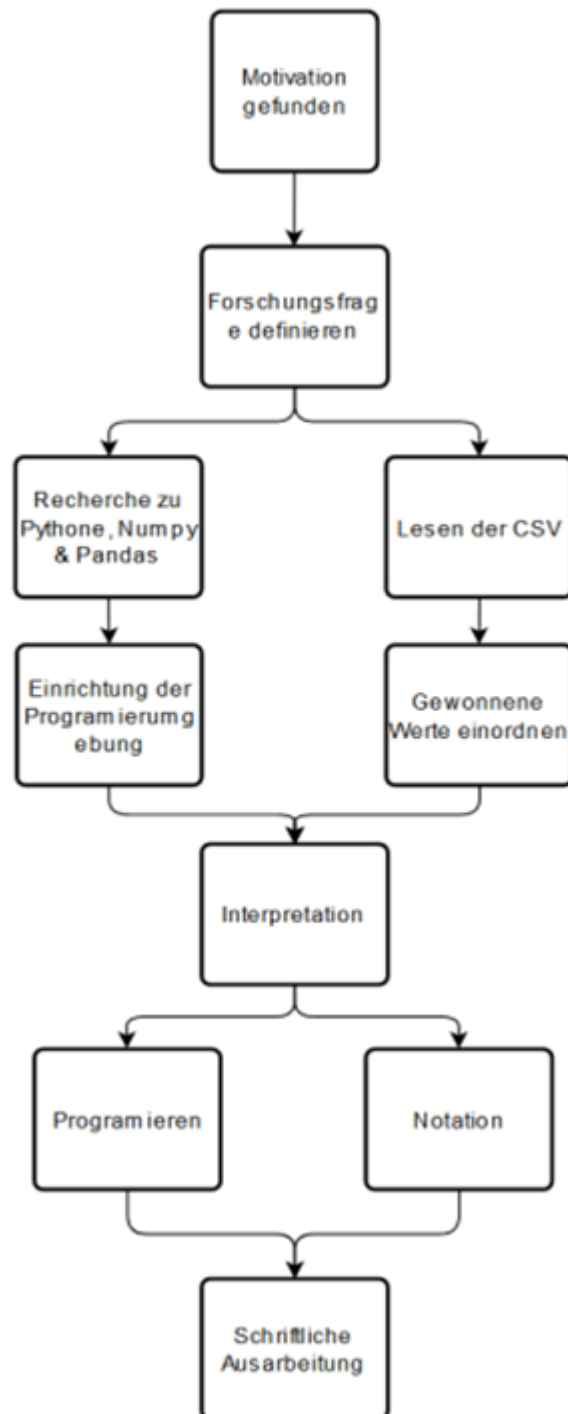
Der praktische Teil dieser Forschung konnte nun beginnen. Das Schreiben unseres Programms verlief aufgrund des Umfangs der Datenmenge mehrmals in gleichen Phasen, um die Datenmenge aufzuteilen für eine klarere Übersicht. Zuerst der Import der Libraries, und anschließend das Lesen der CSV-Dateien aus dem lokalen Speicherort. Mit verschiedenen Methoden, wie z.B. Filtern, Gruppieren oder Dataframing, wurden abschließend die gewünschten Informationen herausgearbeitet.

Nach dem Gewinnen der richtigen Daten, mussten diese noch in einen entsprechenden Kontext eingeordnet werden, um eine Visualisierung und Interpretation zu ermöglichen. Die Visualisierung diente uns im Rahmen dieser Arbeit zur Veranschaulichung der gewonnenen Daten innerhalb derer Einordnungen. Dies empfanden wir als nötig, da der Umfang der Daten aus den CSV-Dateien Missverständnisse ermöglichen zu können schien.

Die Interpretation war der Schritt, in dem alle eingeordneten Daten aus gezielter Ausarbeitung auf Auffälligkeiten überprüft wurden, um eine Aussage zu bilden.

In der Notation wurden alle zugehörigen Ideen und Schlüsse, die durch die Arbeit entstanden sind, festgehalten, um als inhaltliche Vorlage für die Ausarbeitung zu dienen.

Die finale Ausarbeitung ist das genau strukturierte Produkt aller vorherigen Arbeitsschritte, und entstand daher, nachdem alles andere vollzogen war. Der gesamte Ablauf wird nun als Flussdiagramm veranschaulicht:



3.2 Datenbereinigung und Datenbehandlung

Bevor das Programmieren beginnen konnte, musste definiert werden, welche Informationen aus den Datensätzen gewonnen werden sollten. Damit sind Informationen gemeint, die nach der Gewinnung für Aussagen zum Stand der jährlichen Charts und Künstler ausgelegt werden konnten. In unserem Fall legten wir uns auf uns offensichtlich aussagekräftig scheinende Informationen fest, die wenig Interpretationsspielraum ermöglichten. Unsere Datensätze waren jedoch dabei schon ausgelastet, und ermöglichten keine weiteren Datenerhebungen ohne Interpretation oder weitere Forschungsmethoden aus anderen Fachrichtungen. Zuerst wurden „NaNs“, Spalten mit fehlerhaftem oder mangelhaftem Inhalt, mit der Funktion „dropna(axis=0)“ aus beiden Datensätzen herausgefiltert. Diese hätten für Fehler oder andere Störungen gesorgt. Danach wurde der erste Datensatz verwendet, um für eine vereinfachte Übersicht die beliebtesten 20 Künstler der vergangenen vier Jahre zu bestimmen. Dadurch verblieben ausreichende Datenmengen, um Schlüsse ziehen zu können, während ein Überblick bewahrt werden konnte. Dies wurde bewerkstelligt, indem der Datensatz nach „artists“ gruppiert, und mit einer Maske auf die verschiedenen Jahre aufgeteilt wurde. Zusätzlich wurde mit dem Parameter „Global“ gefiltert, sodass nur globale Zahlen, ohne zusätzliche andere Länderzahlen Einfluss nahmen.

Gruppieren nach Künstlern:

```
#Top 20 Künstler in den Charts
df_k1 = df2020.loc[mask]
df_k1_countCountry = df_k1.groupby(["artist"]).count()["count"]
```

Pro Jahr jeweils: Aufteilen nach Jahr mit Maske und Filtern nach dem Parameter „Global“

```
#2020
df2020 = df1
df2020['date'] = pd.to_datetime(df2020['date'])
start_date = '01/01/2020'
end_date = '01/01/2021'
mask = (df2020['date'] > start_date) & (df2020['date'] <= end_date) & (df2020['country'] == "Global")
...
```

Danach wurde die durchschnittliche Länge von Songs kalkuliert. Hierbei wurde der zweite Datensatz, welcher die Songlänge Millisekunden beinhaltet, zur Hilfe gezogen. Hierfür wur-

de eine Liste erstellt, welche die beliebtesten Künstler des ersten Datensatzes beinhaltet. Mit dieser Liste haben wir einen Filter erstellt, der es uns ermöglichte, den ersten Datensatz nach Songs der beliebtesten 20 Künstler zu durchforsten. Daraufhin, haben wir Duplikate mit der Pandas-Funktion „drop_duplicates“ aus dem Datensatz entfernt. Mit diesem gefilterten Datensatz konnten wir den zweiten Datensatz über den gemeinsamen Identifikator „URI“ nach der Länge der Songs filtern. Anschließend wurden weitere Duplikate im zweiten Datensatz entfernt. Danach haben wir mit der Pandas-Funktion „mean()“ den Mittelwert der Länge aller übrigen Songs bestimmen können.

```
'''
~~~~~
Durchschnitt der Songlänge
~~~~~
'''

#Fügt alle Künstler vom Dataframe1 zu einer Liste hinzu
list = df_k1h["artist"].tolist()

#generiert den Filter mithilfe der list
filter = df_k1["artist"].isin(list)
df_k1f = df_k1.loc[filter]
#entfernt Duplikate in df1
df_k1f.drop_duplicates(subset="track", inplace=True, ignore_index=True)

#fügt alle uri von df1 zu einer Liste hinzu
df1_urilist = df_k1f["uri"].tolist()

#erstellt filter und filter direkt df2 nach den passenden Tracks
duration_filter = df2["Uri"].isin(df1_urilist)
df2f = df2.loc[duration_filter]

#entfernt Duplikate in df2
df2f.drop_duplicates(subset="Uri", inplace=True, ignore_index=True)

#gibt Durchschnittslänge aus
df2o = df2f["duration_ms"]
print(df2o.mean())
```

Die nächste behandelte Information ist die Häufigkeit bestimmter Wörter innerhalb der Songtitel. Hierbei verfolgten wir den Ansatz, zuerst die Titel aus Dataframe1 in einem separaten Dataframe zu speichern, welche wir danach nach Duplikaten filterten. Daraufhin wurde mit der Pandas-Funktion „pd.Series“ jeder Songtitel im Dataframe verglichen, und die Wortanzahl der Titel wurde gezählt. Beide Informationen wurden in einer eigenen CSV-Datei für die spätere Auswertung gespeichert.

```

'''
Wörter im Titel check
'''

#filter und fügt den text als variable hinzu
df_text = df_k1["title"]

#entfernt Duplicate
df_k1.drop_duplicates(subset="title", inplace=True, ignore_index=True)

#zählt die am häufigst genutzten wörter und fügt sie zu einer .CSV Datei hinzu
df_wordtitle = pd.Series(' '.join(df_k1['title']).lower().split()).value_counts()[: 142].head(20)
df_wordtitle.to_csv("WordTitle.csv")

```

Das Genre der Songs aus dem eben erwähnten Dataframe war der nächste Anhaltspunkt. Diesen Schritt gingen wir mit den Pandas-Funktionen „groupby“ und „count“, welche es ermöglichten einerseits den Dataframe nach den Genres zu gruppieren, und andererseits die Häufigkeit der Genres zu zählen. Für eine visuelle Veranschaulichung haben wir diese Informationen mit Python und der Programmbibliothek Matplotlib zu einem Kreisdiagramm verarbeitet. Dieses kommt im nächsten Kapitel „Datenanalyse“ vor.

```

'''
Genre der Tracks der beliebtesten 20 Künstler
'''

# Visualisierung von Genre
# Hier werden die Genre aus df2 gezählt sowie auch ihre Namen als Index gespeichert
LsizeGenre = df2f.groupby(["Genre"]).count()["count"]
LLablesGenre = df2f.groupby(["Genre"]).count()["count"].index

# kuchendiagramm wird erstellt und gespeichert
plt.figure(figsize=(7, 7))
plt.pie(LsizeGenre, labels=LLablesGenre, autopct="%1.1f%%")
plt.title("")
plt.savefig("Genre.pdf")
'''

```

Manche Songs haben den sichtbaren Status „Explicit“, um Hörer vorab vor vulgären oder nicht jugendfreien Inhalten zu warnen. Dieser Status wird als absoluter Zustand, vergleichbar mit Boolescher Algebra mit „Wahr“ oder „Falsch“, gespeichert. Hier wurde dieser Zustand für die Songs des Dataframes überprüft. Dies wurde bewerkstelligt, indem wie im letzten Schritt nach Kategorien gruppiert wurde. Dieses Mal war jedoch nicht das Genre, sondern der Status als „Explizit“ oder nicht „Explizit“ entscheidend.

```

'''
Explicit
'''

# Visualisierung von Explicit
# Hier werden die Explicit aus df2 gezählt sowie auch ihre Namen als Index gespeichert
LsizeGenre = df2f.groupby(["Explicit"]).count()["count"]
LLablesGenre = df2f.groupby(["Explicit"]).count()["count"].index

# Figure
plt.figure(figsize=(7, 7))
plt.pie(LsizeGenre, labels=LLablesGenre, autopct="%1.1f%%")
plt.title("")
plt.savefig("Explicit.pdf")

# Just the End
print("Done")

```

Die folgenden Parameter wurden auf dieselbe Weise dem Datensatz entnommen. Daher folgt nun eine Beschreibung der Größen, mit dem geschriebenen zugehörigen Code im Anschluss.

Der Parameter „Danceability“ gibt die Tanzbarkeit eines Songs in Prozent an. Je höher der Wert, desto einfacher lässt sich zu einem Song im Takt tanzen. Dabei wird der Wert als eine Zahl zwischen 0 und 1 angegeben.

„Loudness“ ist der Parameter für die Lautheit eines Songs. Damit ist die Dichte des akustischen Signals gemeint. Je undynamischer die Lautstärke eines Songs ist, desto dichter ist die Lautstärke, wodurch beim Erhöhen der gesamten Lautstärke eine lautere Wirkung erzielt wird, als bei einem Song mit dynamischer Lautstärke. Diesen Wert misst man in „LUFS“, wobei die Werte stets negativ sind. Je näher er dem Wert 0 kommt, desto lauter und dichter klingt ein Song.

Die „Speechiness“ beschreibt die Dichte an Sprache innerhalb eines Songs. Diesen Wert gibt man auch als Zahl zwischen 0 bis 1 an. Je größer der Wert, desto mehr Sprache kommt in einem Song vor.

Der Parameter namens „Acoustics“ beschreibt, wie akustisch, d.h. kabellos ein Song entstanden ist. Diesen Wert gibt man auch als Zahl zwischen 0 bis 1 an. Je größer der Wert, desto akustischer klingt ein Song.

Zuletzt die Geschwindigkeit eines Songs. „Tempo“, gemessen in Schlägen pro Minute, oder auch BPM für „Beats per minute“, wird als natürliche Zahl angegeben.

Jeder der eben genannten Parameter wurde mit derselben Logik aus dem gefilterten Datensatz aus den vorherigen Datenerhebungen gewonnen. Zuerst wurde der gewünschte

Parameter aus dem Datensatz gefiltert. In der Filterfunktion wurde erneut mit „mean()“ gearbeitet, um den direkten Mittelwert des Parameters zu den Songs aus den Charts der 20 beliebtesten Künstler zu erhalten.

```
'''
Danceability
'''
#berechnet den durchschnitt der deancability
df2fdance_string = df2f["danceability"].mean()
print("Danceability: " + str(df2fdance_string))

'''
Loudness
'''
#berechnet die durchschnittliche loudness
df2floud_string = df2f["loudness"].mean()
print("Loudness: " + str(df2floud_string))

'''
Speechiness
'''
#berechnet die durchschnittliche speechiness
df2fspeech_string = df2f["speechiness"].mean()
print("Speechiness: " + str(df2fspeech_string))

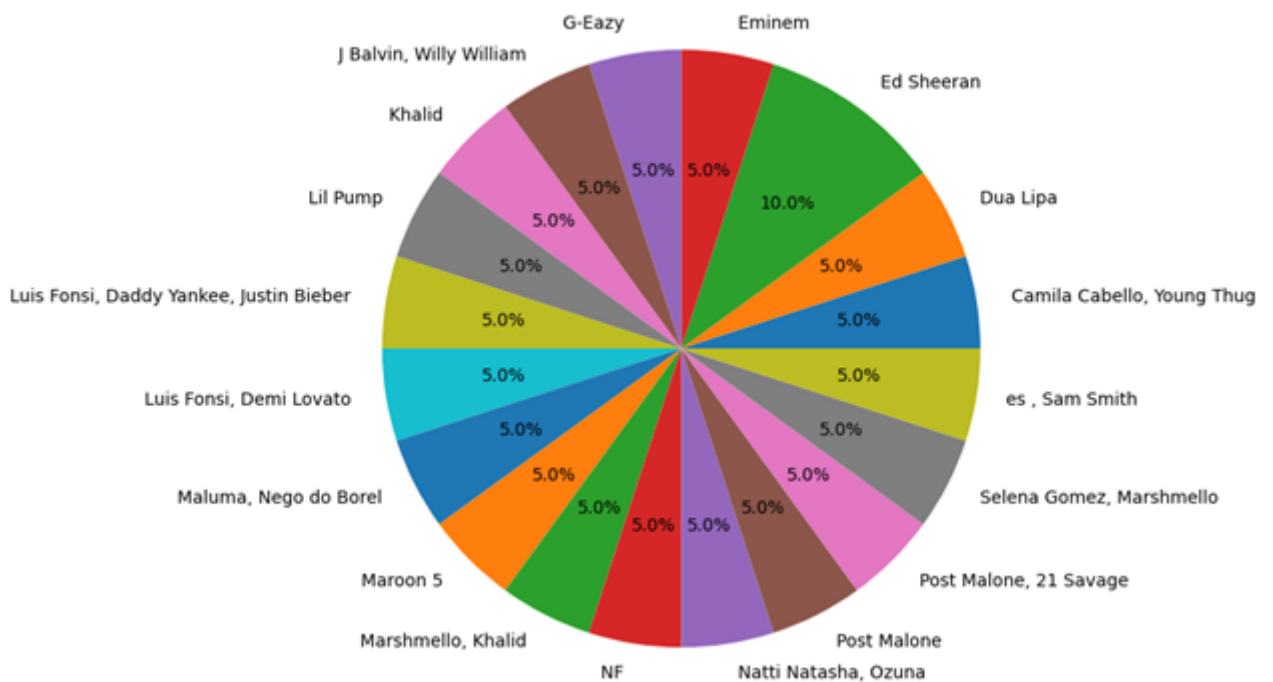
'''
Acoustics
'''
#berechnet die durchschnittlichen Acoustics
df2facoust_string = df2f["acoustics"].mean()
print("Acoustics: " + str(df2facoust_string))

'''
Tempo
'''
#berechnet das durchschnittliche Tempo
df2ftempo_string = df2f["tempo"].mean()
print("Tempo: " + str(df2ftempo_string))
```

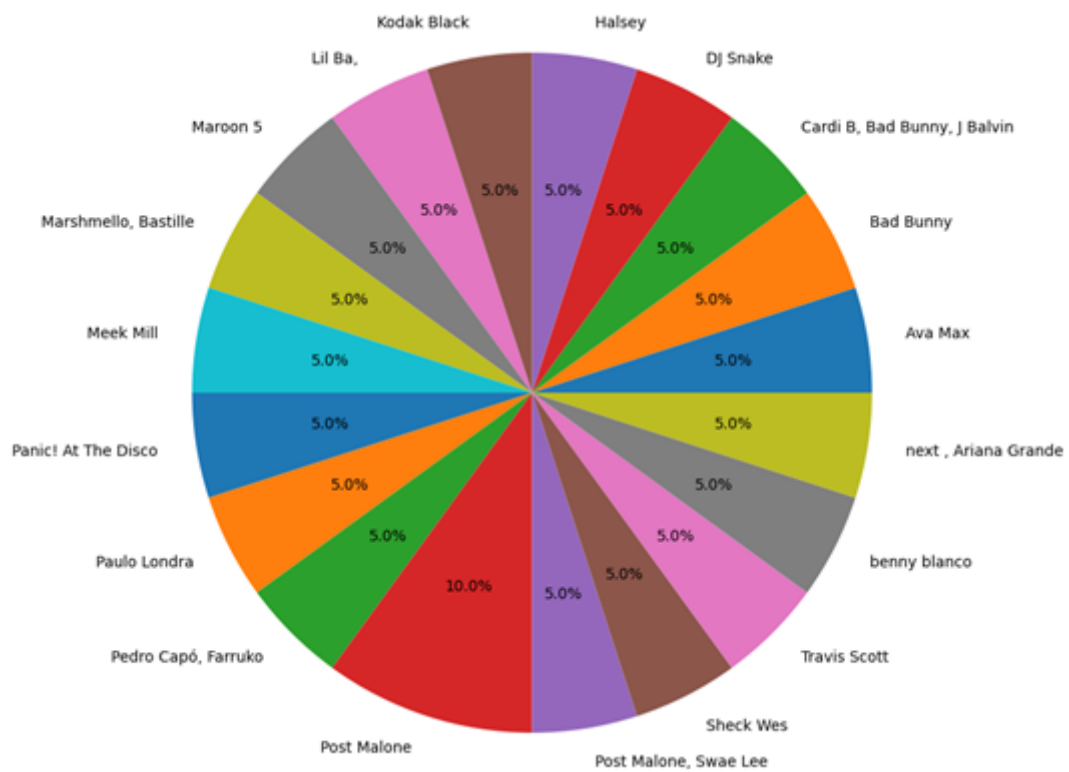
4 Ergebnisse

Nach der Datenbereinigung erhielten wir für jede gewünschte Information ein handfestes Ergebnis. Die Analyse der 20 beliebtesten Künstler auf Spotify der Jahre 2017 bis 2020 ergab folgende Ergebnisse, wobei die Prozentzahlen den Anteil an den gesamten Aufrufen der Charts angeben:

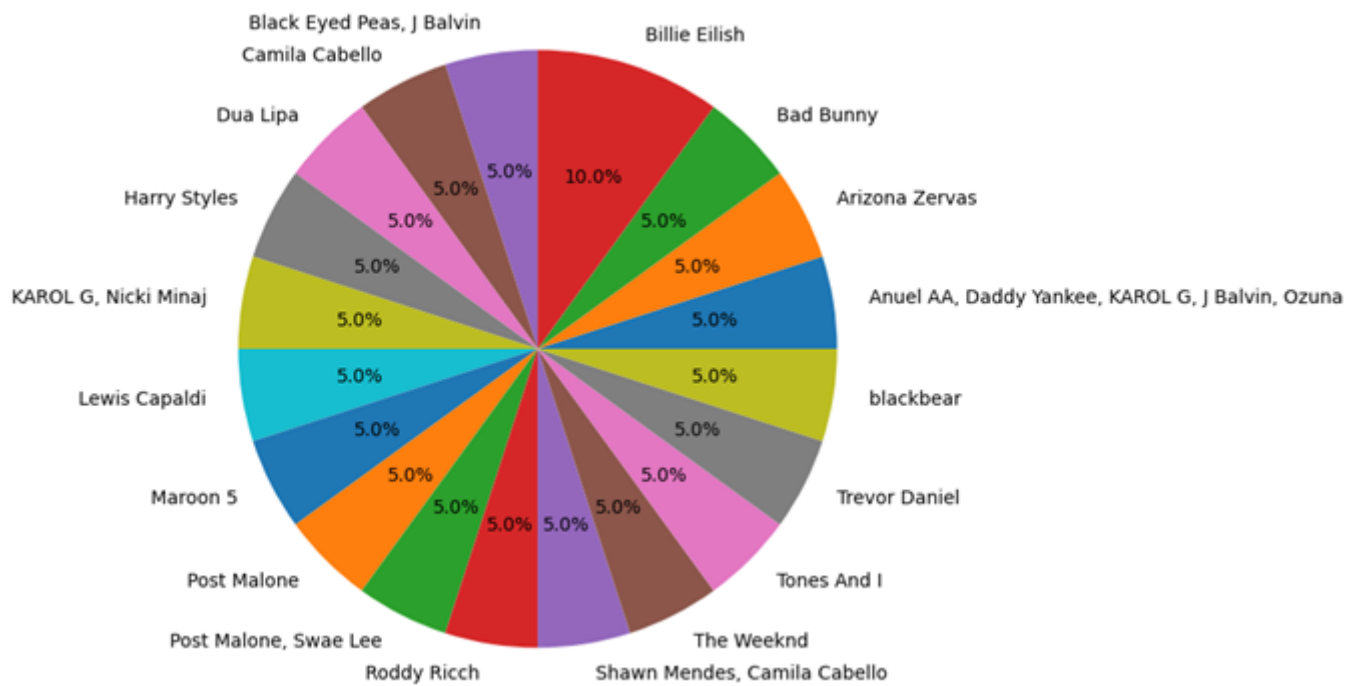
2017



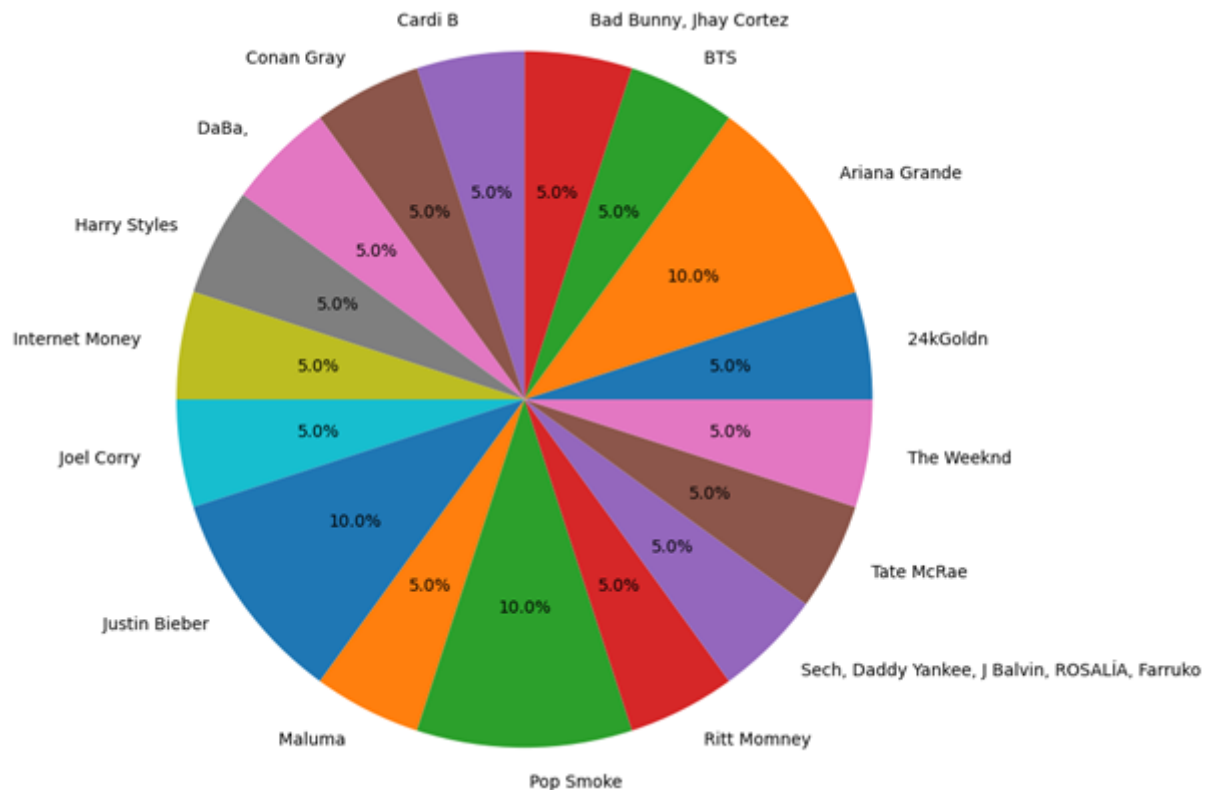
2018



2019



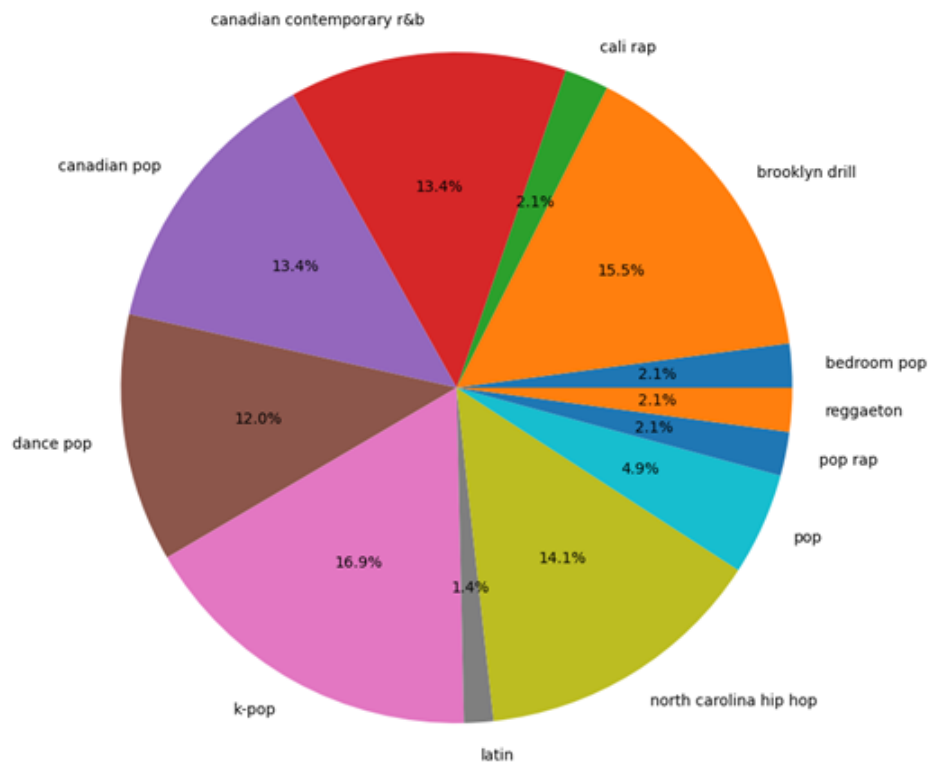
2020



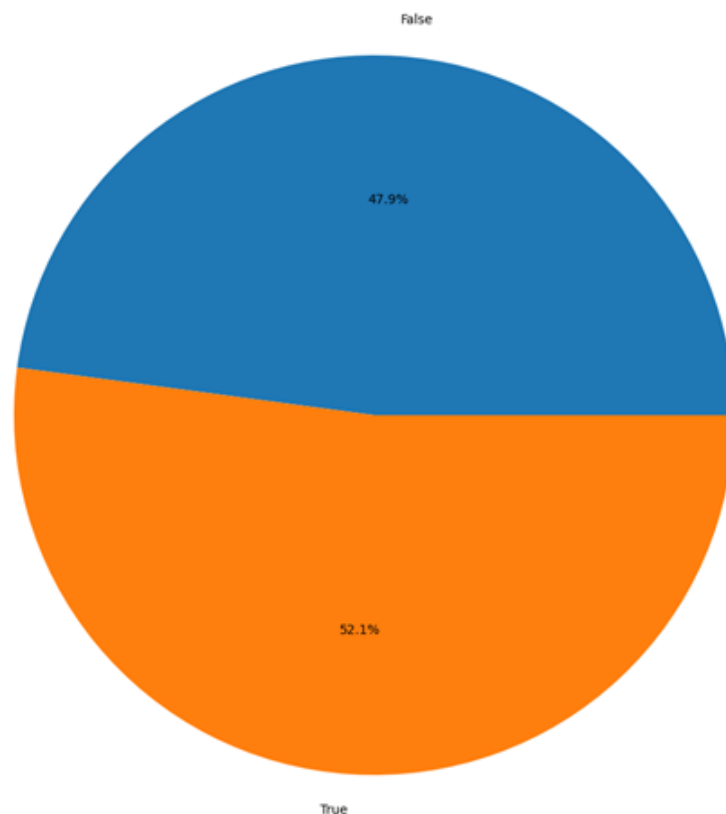
Unsere Untersuchung der durchschnittlichen Längen der Songs der 20 beliebtesten Künstler des Jahres 2020 ergab einen Wert von 3,204 Minuten. Das entspricht drei Minuten und zwölf Sekunden.

Die Suche in unserer eigenen CSV-Datei nach Auffälligkeiten bei der Wortwahl von Songtiteln führte zum Ergebnis, dass die fünf Worte „the“, „me“, „you“, „remix“ und „my“ am häufigsten in Songtiteln der 20 beliebtesten Künstler der Charts im Jahr 2020 auftraten. Wobei „the“ 75 mal, „me“ 63 mal, „you“ 59 mal, „remix“ 40 mal und „my“ 38 vorkamen.

Die Analyse zur Beliebtheit bestimmter Genres führte zu einem mehrdeutigen Ergebnis. D.h. mehrere Musikrichtungen teilen sich den Status als beliebtestes Genre. Mit folgender Verteilung traten diverse Musikrichtungen in den Charts 2020 auf:



Aus persönlichem Interesse untersuchten wir auch den Verhalt von als explizit eingestuften Songs. Hier wurde wie zuvor beschrieben die Häufigkeit von Songs mit diesem Status in den Charts überprüft. Dabei entstand folgendes Ergebnis, wobei Orange für den Anteil von expliziten Titeln steht:



Mit einer durchschnittlichen „Danceability“ von ca. 0,67 sind die Charts des Jahres 2020 eindeutig für das Tanzen geeignet.

Die durchschnittliche „Loudness“ der Songs ergab einen Wert von ca. -6,3 LUFS. Es scheint also eine Neigung zu eher laut klingenden Songs zu geben.

Die „Speechiness“ war dabei im Schnitt weniger hoch. Ihr Wert lag im Durchschnitt bei ca. 0,145. Daraus folgt, dass die meisten Songs der Charts unserer Künstler des Jahres 2020 zwar Sprache in ihren Songs nutzen, jedoch nicht ausgiebig.

Der Parameter „Acoustics“ ergab einen durchschnittlichen Wert von ca. 0,2. Damit bestehen alle Songs unserer Künstler im Durchschnitt zu einem Fünftel aus akustisch erzeugten Tönen. Das deutet auf eine Neigung zu elektronisch erzeugten Tönen hin.

Zuletzt ergab die Geschwindigkeit der Songs einen durchschnittlichen Wert von ca. 125 BPM. Dieser Wert passt zu der oben genannten durchschnittlichen Tanzbarkeit.

5 Limitation

Da jede Arbeit in den Kontext von dessen ursprünglichen Bedingungen einzuordnen ist, kann eine Veränderung dieser Bedingungen für eine Abweichung der Aussagekraft sorgen. Durch das Aufzeigen der Ursprungsbedingungen, werden ebenfalls die Grenzen der Aussagekraft aufgezeigt. Auch diese Arbeit verfügt über Grenzen, welche nun nacheinander aufgeführt werden.

Eine Limitation dieser Arbeit ist die individuelle Fähigkeit im Programmieren und in der Datenanalyse. Die eigene Voreingenommenheit was die Gewichtung von Argumenten oder der allgemeinen Arbeitsweise betrifft hat Einfluss auf den gesamten Verlauf dieser Arbeit.

Eine weitere Eingrenzung ist der Datensatz selbst. Ein vorgegebener Datensatz beinhaltet genau festgelegte Zeiträume, Orte, Personen und andere Parameter. Die Verbindlichkeit dieser Informationen kann nicht ohne weitere Datenerhebung überprüft werden. Außerdem können Informationen veraltet oder falsch sein. Die Größe des Datensatzes ist ebenso eine Eingrenzung, wobei eine größere Datenmenge für mehr Daten und somit weniger Limitation sorgt.

Die Forschungsfrage selbst bildet eine weitere Limitation, denn hier wird das Ziel der Arbeit festgelegt. In unserem Fall wurden andere mögliche Fragen ausgeschlossen, um für die Einhaltung der Vorgaben dieser Arbeit eine einzelne Frage zu beantworten. Da die Forschungsfrage zusätzlich Rahmenbedingungen schafft, wie in unserem Fall die Festlegung auf Daten aus Spotify, wird die Arbeit noch weiter eingegrenzt. Spotify ist eine einzelne Plattform, doch die Frage verlangt eine Aussage über allgemeine Musikrends. Für eine höhere Aussagekraft müssten Daten von weiteren Plattformen analysiert werden.

Zusätzlich ist die Entscheidung über die Wahl der zu untersuchenden Kategorien an Informationen eine ausschlaggebende Begrenzung. Um die Übersicht zu bewahren, wurden bewusst nicht alle Kategorien der gegebenen Parameter analysiert. Parameter mit Angabe zu z.B. der Emotion eines Songs, oder der Beliebtheit wurden ignoriert. Außerdem existieren

Parameter außerhalb unserer Datensätze wie z.B. die „Skiprate“, welche angibt in welcher Rate Songs innerhalb der ersten wenigen Sekunden nach dem Start übersprungen werden. Es existieren noch weitere Parameter, die die Aussagekraft dieser Arbeit verbessern könnten. Doch damit wäre der Umfang dieser Arbeit zu groß, weshalb bestimmte Limitationen in unserem Fall von Vorteil sind.

6 Fazit

Nachdem nun die Ergebnisse erhalten wurden, können mit einem Interpretationsspielraum Aussagen oder Vermutungen aufgestellt werden. Die Verteilung der Künstler in den Charts der Jahre 2017 bis 2020 zeigt eine auffällige Gleichheit der Anteile der meisten Künstler. In jedem Jahr haben alle bis einzelne andere Künstler denselben Anteil an Chartplatzierungen. Außerdem bleiben viele der Künstler konstant beliebt in diesem Zeitraum, während andere ein plötzliches Eintreffen oder auch Verlassen der Charts aufweisen. Das alles deutet auf eine Verfolgung von Trends von bereits etablierten Künstlern hin. Während einige Künstler Trends setzen bzw. Trends aktiv verfolgen, um relevant zu sein, werden erzielen andere weniger Chartplatzierungen und fallen möglicherweise vollständig aus den Charts heraus.

Die durchschnittliche Songlänge von 3 Minuten und 12 Sekunden erwies sich als kurz, verglichen mit Erfahrungswerten. Dies spricht für einen Trend, der mit der durchschnittlichen Aufmerksamkeitsspanne der Konsumenten zusammenhängen könnte. Derartige Trends sind schließlich Produkte der Nachfrage.

Die fünf Worte „the“, „me“, „you“, „remix“ und „my“ als häufigste Wortwahl der Songtitel ergaben keine auffälligen Schlüsse, da es sich bei diesen Worten um allgemein gängiges Vokabular bzw. um das Wort „remix“, aus der elektronischen Musik für eine Neuauflage, handelt. Hierbei hätte ein gezielteres Filtern nach z.B. der Wortlänge für ein aussagekräftigeres Ergebnis gesorgt.

Die beliebtesten Genres wiesen eine Auffälligkeit auf unter Pop-Musik und Musik aus der Hip-Hop-Kultur. Es kommen ausschließlich jene Genres in den Charts vor, jedoch als Kreuzungen mit anderen Genres wie z.B. Dance.

Über die Hälfte der Songs aus den Charts gelten als „Explicit“, also nicht kinderfreundlich. Eine mögliche Begründung könnte die Zielgruppe von moderner Musik sein, welche typischerweise junge Menschen sind. Wobei junge Menschen rebellisch sein möchten.

Die Werte von „Danceability“, „Loudness“, „Speechiness“, „Acoustics“ und „Tempo“ deuten zusammen mit den vorherigen Informationen auf eine Neigung zu elektronischen und tanzbaren Einflüssen auf die aktuelle Pop-Musik, jedoch ohne ausführlichen Einsatz von Sprache. Dieser ist trotzdem nicht unwesentlich, sondern ein fester Teil der Formel.

All diese Parameter sind bewusst von den Künstlern getroffen, da es eindeutige Trends gibt. Das bedeutet, dass in Hinblick auf die geprüften Werte, die erhaltenen Ergebnisse als Orientierung oder auch „Erfolgsfaktoren“ betrachtet werden können.

Die Zukunft wird möglicherweise eine weitere Steigerung in dieser Entwicklung mit sich bringen, wodurch beispielsweise elektronische Einflüsse häufiger vorkommen könnten. Die Neigungen dazu sprechen für einen fortlaufenden derartigen Verlauf, und nicht für eine Veränderung oder Innovation der Chart-Musik.

7 Literaturverzeichnis

Dörn, Sebastian (2020): Python lernen in abgeschlossenen Lerneinheiten. Programmieren für Einsteiger mit vielen Beispielen. Wiesbaden, Heidelberg: Springer Vieweg (Lehrbuch).

Gomes, Inês; Pereira, Inês; Soares, Inês; Antunes, Mariana; Au-Yong-Oliveira, Manuel (2021): Keeping the Beat on: A Case Study of Spotify. In: Álvaro Rocha, Hojjat Adeli, Gintautas Dzemyda, Fernando Moreira und Ana Maria Ramalho Correia (Hg.): Trends and Applications in Information Systems and Technologies, Bd. 1366. Cham: Springer International Publishing (Advances in Intelligent Systems and Computing), S. 337–352.

Rocha, Álvaro; Adeli, Hojjat; Dzemyda, Gintautas; Moreira, Fernando; Ramalho Correia, Ana Maria (Hg.) (2021): Trends and Applications in Information Systems and Technologies. Cham: Springer International Publishing (Advances in Intelligent Systems and Computing).

8 Ehrenwörtliche Erklärung

„Wir versichern, dass die vorliegende Arbeit von uns selbständig und ausschließlich unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt wurde. Alle Stellen, die wörtlich oder annähernd aus Veröffentlichungen entnommen sind, haben wir als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form, auch nicht in Teilen, keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.“

Ort, Datum

Beilstein, 13.01.22

Unterschrift

Avram Ivanovic

Ort, Datum

Stuttgart, 13.01.22

Unterschrift

David Pehar

Ort, Datum

Sulzfeld, 13.01.22

Unterschrift

Suphi Pembe