

Trends in Western Popular Music: Analyzing Top Album Charts and Online User Opinions

UNIVERSITY OF TURKU
Department of Computing
Master of Science Thesis
Computer Science
May 2024
Heikki Vuorinen

Supervisors:
Samuli Laato
Paavo Nevalainen

UNIVERSITY OF TURKU
Department of Computing

HEIKKI VUORINEN: Trends in Western Popular Music: Analyzing Top Album Charts
and Online User Opinions

Master of Science Thesis, 49 p., 4 app. p.
Computer Science
May 2024

During recent years, music has become an increasingly larger part of our lives due to the vast availability of music online. Technological innovations have always played a key role in the music industry, changing the ways music is recorded, distributed and consumed. However, not everyone agrees that these changes have improved the quality of music and it's quite common among music enthusiasts to hear the phrase "Music was better before". Of course, the perception of music is highly subjective, but it's possible to examine the changes that could lead consumers to have this view. In this thesis our goal is to examine how the technological evolution of music distribution has affected the trends and dynamics of music consumption, based on historical Billboard 200 charts, which represent the weekly 200 best selling albums in the US. Our goal was to evaluate topics such as the amount of yearly unique albums, the popularity of genres and the current Rate Your Music user ratings of past and contemporary albums. We found that the popularity of the CD format led to an increase in unique album appearances in the list up until around the time streaming started gaining popularity. The popularity of streaming also led to an increase in the sales of older records, which could now be monetized through streaming services. Additionally, we visualized the genre popularity over time and examined the online user based album ratings regarding time and genres. We found that the overall average user ratings decreased during the late 1990s, and that there are significant differences in the way ratings are distributed accross genres, most notably rock being the highest rated with an average of 3.39 stars on a scale from 1 to 5 stars, and hip hop the lowest with 2.90/5 stars average.

Keywords: Billboard 200, Popular Music, Music Trends, Music Industry Dynamics,
Rate Your Music, Music Genre Trends, Music Streaming

Table of Contents

1	Introduction	1
1.1	Structure of the thesis	3
2	Background and Literature Review	4
2.1	Technological impact on the music industry	4
2.1.1	The era of physical recording media	4
2.1.2	Digital recordings and the impact of the internet	5
2.1.3	Impact of streaming in listening habits	6
2.2	Perception of old vs contemporary music	7
2.3	Genre trends and popularity over time	8
2.4	Genre and automatic genre classification	10
3	Materials and Methods	12
3.1	Data sources and collection	12
3.1.1	Data sources	13
3.1.2	Data collection and variables	14
3.1.3	Web scraping ethics	17
3.1.4	Data cleanup and validation	18
3.2	Analysis	19
3.2.1	Identifying unique albums and artists	19
3.2.2	Distribution of chartlisted weeks among albums and artists . .	20
3.2.3	Genre analysis	21
3.2.4	Analyzing RYM ratings	22
4	Results and Findings	23
4.1	Results of Billboard analysis	23

4.1.1	Yearly unique albums	23
4.1.2	Most Successful Acts	25
4.1.3	Distribution of chart presence	27
4.2	Genre proportions of charting albums over time	30
4.2.1	Overall genre distribution	30
4.2.2	Genre distribution over time	31
4.3	User ratings	32
5	Discussion	38
5.1	The variety of yearly unique album appearances on the Billboard 200 chart	38
5.1.1	The return of previous hit records	39
5.2	The most succesful acts and the distribution of chart presence among artists	41
5.3	Most prominent genres accross the history and the value of genre as a music classification tool	42
5.4	User opinions in Rate Your Music and the percieved quality of music accross different decades and genres	44
5.5	Theoretical and practical contributions	46
5.6	Limitations and future work	46
6	Conclusion	48
	References	50
	Appendices	
A		A-1

1 Introduction

Music plays a large part in our everyday lives. In the 2010s, music streaming services have become the most common way of consuming music, allowing users to listen to a vast catalogue of music almost everywhere we go [1], [2]. According to a report by IFPI published in 2018, the average consumer listens to 17.8 hours of music each week with 75% of individuals listening on their smartphones and 86% using on-demand streaming services such as YouTube or Spotify. For example, music is commonly listened to on commutes, while working and while relaxing at home. [1]

The ubiquitous availability of music is a relatively recent phenomenon. While the availability of music began to rise in the early 1990s when compact disks started gaining popularity, it was the internet that truly revolutionized how music is distributed and consumed. Even before the rise of streaming services, the internet allowed users to share and download music, enabling new ways of discovering new music. [3]–[5]

Advancements in technology have not only transformed the way we consume music but also how it is written, produced and recorded. Digital audio workstations or DAWs have made the process of creating music possible without the access to any physical instruments or expensive studio equipment, making music production more accessible to everyone. The availability of home computers powerful enough to produce and record music at home, coupled with the development of DAWs and the ability to self-publish music on the internet, has allowed independent artists to gain

a significant following online, with some achieving mainstream success. Additionally, the wealth of online resources and guides has democratized the process of recording and producing music, making it more accessible to aspiring musicians. [3]

With the democratization of music production and publishing caused by technological advancement and the internet, not only the availability, but also the amount of published music has increased. This possibility has led to a wider variability in music, giving birth to various new subgenres [6]. However, the advances in music production and distribution capabilities have not *prima facie* translated into better and more enjoyable music. Of course, the perceived quality of music is highly subjective, but it's quite common among music enthusiasts to hear that older music is "better", especially when talking about popular music [7]–[9]. This perception could depend on various factors and biases, such as associating memories with certain records, or the fact that we've had time to hear the influences that some certain records have left on popular music. This is inherently something that we cannot know for sure about contemporary music.

The goal of this thesis is to explore the topic of change in popular music and find any patterns or trends that might help us understand how the production and consumption of music has changed over the last few decades. We analyze historical music trends to gain an understanding of the current musical landscape and what has led into it. This involves examining historical chart data to identify shifts in popular music. We also examine the perceived quality of contemporary popular music when compared to past popular music. Similar topics have been studied before, but as the music landscape constantly evolves, our goal is to provide additional results regarding the most recent years to underline the impact of online music distribution and streaming services. Additionally, our goal is to tie our results to online user ratings, as this hasn't been studied before regarding music. We have defined the research questions of this thesis as follows:

- RQ1: How is the Billboard 200 album chart presence distributed among various artists and genres and what genres have been the most prominent over the last few decades?
- RQ2: Has there been shifts in the Billboard 200 chart dynamics influenced by recent technological changes in music distribution?
- RQ4: How online user reviews in Rate Your Music are distributed among various genres and contemporary vs past popular music?

1.1 Structure of the thesis

The thesis is structured as follows; in the next chapter we will provide background for this thesis and a comprehensive literature review, providing insight into what previous work has already been conducted in this area. In Chapter 3 we will look into the data collection process and methods used in the data analyses conducted in this study. The chapter contains a comprehensive description of how the data sets were collected, and how we have utilized them to answer our research questions. In Chapter 4 we will explain and visualize the results of our analysis. In Chapter 5, we will discuss the results comparing them to past studies, and aim to draw any conclusions based on the results of this thesis. We will also address any limitations of this study and discuss possible improvements and any uncovered areas for future research. Lastly, in Chapter 6, we will summarize the results of this thesis.

2 Background and Literature Review

2.1 Technological impact on the music industry

In this section, we explore the historical evolution of the music industry, focusing on key technological advancements that have shaped it into its current form. By examining these developments, we gain insight into how music distribution and consumption have been influenced by technology over time. Technological developments have always been in the center of the development of the music industry, as it was the invention of the first recording devices that first led to the birth of the recording industry, nowadays often used synonymously with the music industry [3], [10].

2.1.1 The era of physical recording media

In the early 20th century, radio and vinyl recordings were the common ways of distributing music. Since the birth of radio and vinyl, during the 1980s it was the compact cassette that again revolutionized the way music was consumed [11]. Cassette tapes allowed consumers to record and share music affordably, and the portable Sony Walkman cassette player allowed consumers to listen to music anywhere they go. [3]

Soon after the compact cassette, the compact disk (CD), was invented. CDs

retained the portability of the cassette, but had many advantages over both cassette tapes and vinyl disks. The audio quality of CDs doesn't degrade as a result of playbacks, and arguably, CDs offered better quality overall [3]. They were the first popular recording medium that stored the recordings in a digital format. The recordings are stored using a sample rate of 44.1kHz, and according to the Nyquist-Shannon sampling theorem, this results in a bandwidth of 22.05kHz, corresponding to the hearing limit of most humans [3], [12].

2.1.2 Digital recordings and the impact of the internet

Since the late 1990s, MP3 and other digital formats emerged as the most common medium for music distribution. Combined with the development of the internet, consumers could now find and share music online. Apple's iTunes became the most dominant marketplace to buy and download licensed music, which could then be listened to on computers or portable mp3 players. During each time a new format has been popularized, it has boosted sales due to the so-called album-replacement cycle, which means consumers replacing their old records with ones in the new format. However, a large part of digital music was also shared illegally, resulting in lost sales and revenue for the recording labels and artists. [3], [4]

The transition from physical media to digital files also resulted in a fundamental change to the concept of music ownership. Rather than owning physical copies of the recordings, music ownership now meant owning a license to use the music. Streaming platforms, most notably Spotify, have been built on this concept. The idea behind music streaming is that users can listen to an unlimited catalog of music, while paying for a monthly subscription. However, this means that none of the music is actually owned by the users, and the access to music is lost in case the subscription is terminated. [3]–[5]

2.1.3 Impact of streaming in listening habits

Though the wide range adoption of streaming services has resulted in lowering album sales, it has allowed for a constant stream of income, since users have to keep paying for the services [13]. Music streaming has also again fundamentally shifted the way listeners consume music. Since there is practically an endless library of music, listeners can access an unlimited variety of music without any extra cost. Researchers have shown that the streaming has increased the variety of music people listen to. In just first two weeks after the adoption of Spotify, the number of unique artists increased by 62%, the number of unique songs by 49% and the number of unique genres by 43% [5].

In the days of physical media, consumers were mostly forced to buy entire albums which led to them listening to the albums as one coherent piece. Streaming has changed listening habits in a way that most users now listen to particular songs and combine them into playlists, which has led to them not recognizing album as the actual holistic work of art, but rather the focus is on single songs. [3]

In addition to user-generated playlists, streaming platforms have also incorporated algorithmically generated playlists based on the users' listening habits. One such playlist is the 'Discover Weekly' playlist by Spotify, which is generated based on playlists created by other users and the 'Taste Profile' of each user. The goal of this playlist is to expose users to new music that they haven't heard, but are likely to enjoy. [14].

As a conclusion, it's widely acknowledged that the evolution of technology is fundamental to the evolution of music distribution and consumption. Most notably in the 2010s, music streaming has revolutionized the way most listen and discover new music. Instead of traditional methods such as browsing records at a store, new music is now recommended by an algorithm tailored to meet the user's preferences.

2.2 Perception of old vs contemporary music

As a part of this study focuses on the current listener opinions on music of the past five decades, we want to cover any important factors that affect our perception of older music compared to newer music. Music is believed to have a strong connection to our memory, which is why certain songs can evoke vivid memories from our past [15]. Contemporary music is not tied to our past, which is why older records might inherently feel more enjoyable and meaningful, evoking emotions and memories in a way that is not possible for contemporary music.

Previous studies have examined the familiarity of music from different decades between younger and older participants. Older participants recalled more songs from their youth, but the songs also evoked more memories and stronger emotions than more recent music [16]. The feeling of nostalgia has also been studied regarding video game music, where researchers found that a large portion of YouTube comments on old video game soundtracks were related to feelings of nostalgia and comfort [17].

The feeling of nostalgia can be related to both positive and negative feelings when listening to music. Researchers have found that sadness was the most common feeling related to songs that evoked personal memories, while unfamiliar or disliked songs evoked feelings of irritation. The tendency to experience music evoked nostalgia was also noted to possibly be tied to personality and current mood. [15]

Among other demographic features such as gender, ethnic background and social class, age has been found to be the strongest predictor of musical taste [18]. People tend to prefer music from their late teens to early adulthood, with studies suggesting ages between 17 and 23 years [18], [19]. Studies also have found the relationship between age and music preference to be an inverted U shape, suggesting that people in general enjoy both older and newer music less than music from their preferred age [18].

Based on previous studies, its apparent that age is a major factor in music

preference, and music evoked memories and nostalgia can significantly affect the enjoyment of music. The results suggest that this might be a possible reason why it's quite common to hear people downplaying contemporary popular music, saying that popular music was superior when they were young. However, it is important to note that such claims are largely subjective and based in personal experience, rather than objective analysis.

2.3 Genre trends and popularity over time

Genre is the most common descriptor of music, generally covering musical features such as tempo, rhythm and the instruments used. Over the past few decades, various different genres have risen to prominence, reflecting the evolution of cultural trends and shifts in stylistic preferences in music. Understanding genre preferences helps music publishers shape their business strategies to fit the current musical landscape.

Past researchers have been interested in modeling these musical preferences and how they reflect the current cultural environment. It has been shown that an individual's personality, personal values and cultural values affect the preference in music [20].

Petitbon & Hitchcock [21] modeled genre appearances in the *Billboard Hot 100* charts and *Pazz & Jop* critic polls. They also used various socio-economic factors such as unemployment rate, political environment and gross domestic product to model how these variables affected the genre preferences. The predictors were very successful in predicting genre appearances in the *Billboard Hot 100*, and slightly less successful in predicting the critic polls. The socio-economic factors were found suitable for predicting the trends, which implies that they have a significant impact on the musical preferences of the masses. However, the limitation of their models was that they could only predict one year into the future. Another concern noted by the authors of this study was that the assignment of genres was complex in cases

where the titles crossed over various different styles.

In 2016, The DataFace, a data visualization company, published an article showing how they studied the genre popularity over time using the *Billboard Hot 100* chart as a source. Primarily, they utilized the Spotify API to fetch genres for each single, successfully categorizing 79% of the songs. The rest of the data was filled from EveryNoise¹, AllMusic², and Wikipedia. They also noted challenges during the genre classification, due to Spotify classifying artists into over 1300 different genres total. They ended up using a two way approach, creating a mapping of Spotify genres into broader "overarching" genres, such as "rock" or "soul". As an end result, they ended up using 16 of these different major genres. Their visualization shows that historically, rock has been the most prominent genre, with R&B gaining popularity in the mid-1990s, and pop becoming the most prominent genre since the mid-2000s. [22]

Mauch et. Al [23] combined the use of Last FM tags and music information retrieval methods to examine how various musical features have evolved in the Billboard Hot 100 singles over time. Based on 30 second samples of the song, they analyzed them for pre-defined 'topics' such as energy, major/minor chords, 7th chords and percussion. Based on these topics the songs were then clustered using k-means clustering, and the most common Last FM tags were assigned to each cluster. The researchers also looked into musical diversity in the charts and based off of stylistic changes, they identified three major musical 'revolutions', during 1964, 1983 and 1991 where the stylistic changes were the most rapid.

¹<https://www.everynoise.com>

²<https://www.allmusic.com>

2.4 Genre and automatic genre classification

Genre is probably the most important descriptor of music, used by both publishers and the consumers. However, there is no consistent taxonomy and genre is seen as a relatively subjective matter. The upbringing of online music distribution platforms has brought the need for a consistent genre taxonomy of musical titles. Pachet & Cazaly [24] compared genre taxonomies from three different internet sources, Amazon, AllMusic and mp3.com, and found that they shared only 70 genres in common. They also proposed their own taxonomy, with the goal of being objective and consistent.

Aucouturier & Pachet [25] described genre as an ill-defined notion, because it is not an intrinsic notion defined by the content itself, but rather an extrinsic notion defined by external factors such as the time and place the music was recorded. They also criticized past attempts to create objective taxonomies due to the constantly evolving music landscape and the overall challenges in describing genres objectively, which have hindered such efforts.

The challenges in classifying music manually have led to an interest in automated genre classification based in feature extraction using music information retrieval (MIR) methods where features such as tempo, timbre and rhythm are extracted from the musical content itself. Early works in automatic genre classification use feature extraction combined with supervised learning methods such as the K-nearest neighbours classifier and gaussian mixture model [26].

The success of Deep Neural Networks in the 2010s has inspired researchers to apply these methods to genre classification. Representing music in a vector form required by neural networks is not a straightforward task due to the high and varying sample rates of sound files. One approach to solving this problem are spectrograms which are 2D image representations of the sound files that can be fed to convolutional neural network, thus eliminating the need for feature extraction, letting the network

itself learn the essential features required for the classification task [27].

Feature engineering based methods and spectrogram based CNN methods have been compared, with feature extraction models achieving 59% accuracy with a XG-Boost classifier, and CNN methods achieving 64% accuracy. The best accuracy (65%) was achieved using ensemble learning method combining these two classifiers [28].

These content based automatic genre classifiers have been compared to human classification, showing that humans can achieve a 90% accuracy when classifying music into genres, while automatic genre classification achieved a 65% accuracy on the same dataset [29]. This shows that genre classification is very subjective, and automatic genre classification still has a long way to achieve the same level of accuracy as humans. As previously mentioned, genre is also not something that can be entirely inherited from the musical content. Genre as we understand it, is also dependent on external factors, and consequently, relying solely on musical features for genre classification overlooks crucial contextual elements, limiting the effectiveness of automatic genre classification methods.

3 Materials and Methods

3.1 Data sources and collection

To answer the research questions presented in this thesis, we needed to collect some form of historical data reflecting the evolution of music consumption over the years. Possible sources for such data were Billboard and Official Charts. Billboard is an American music magazine which has been compiling album sales in the US into charts for many decades. Official Charts, previously Official UK Charts Company, provides charts based on the sales in the UK.

Billboard was chosen as the most optimal source, due to the US market being larger and thus representing a larger amount of consumers. Billboard also has a reputation as the music industry standard for chart data, and is also used by many professionals as their primary source. The Billboard website also provides open access to all of the historical charts.

In addition to Billboard data, expanding the dataset with additional album metadata was essential for more extensive analysis. To accomplish this, we considered multiple possible options, such as the APIs provided by Spotify and Last FM, which has been used by previous researchers, and community based platforms such as Album Of The Year (AOTY), Rate Your Music (RYM) and AllMusic.

Spotify as the largest music streaming platform was considered, but they denied the use of their API for this thesis. Last FM does not categorize albums in a strict

genre hierarchy, but rather uses user-defined tags, which were not optimal for this research. Rate Your Music has its own taxonomy of genres, while also providing other metadata such as the language. In addition, the site has the ability for any user to rate and review the album, which could be used to measure current music enthusiast views on past and contemporary popular music. RYM was chosen over alternatives such as AOTY and AllMusic due to generally having a significantly larger amount of reviews, as shown in Table 3.1.

Album	RYM	AOTY	AllMusic
Pink Floyd - The Dark Side of the Moon (1973)	78 729	8 843	13 680
Taylor Swift - Midnights (2022)	10 507	5 465	349
The Beatles - Abbey Road (1969)	61 332	6 819	11 932
Metallica - Metallica (1991)	29 450	2 844	5 574
Radiohead - In Rainbows (2007)	77 305	14 616	6 765

Table 3.1: Amount of user reviews for given albums [30]–[32]

3.1.1 Data sources

Billboard

Historical chart data was collected from Billboard.com, an American magazine focused on music and entertainment. Billboard magazine is well known for its charts such as the Billboard Hot 100 and Billboard 200, which are one of the oldest ongoing charts published by Billboard, both introduced in the late 1950’s. [33]

Billboard charts are updated weekly, reflecting changes in popular music trends and shifts in genre popularity. Since music consumption has changed significantly since the introduction of various digital platforms and streaming services, both charts have also been updated to include streamed plays and digital sales [34], [35]. These charts only track music consumption in the United States, but as the US

music industry represents the greatest share of the western world, they can also be thought to generally represent trends in the western world [36].

Billboard charts function as a suitable basis for the data set used in this thesis, as they have covered music consumption since the beginning of the ‘album era’, a term used to describe the timeframe where full length albums have been the most common way of publishing and consuming music [37]. The billboard charts are also commonly referred to as the music industry standard for measuring the commercial success of music.

Rate Your Music

Rate Your Music (RYM) is a community driven online platform where music enthusiasts can catalog, rate and review albums. RYM was launched in the 2000’s and as users can contribute missing records and information to the service, it has since grown into a comprehensive database of music metadata.

As a music review platform, Rate Your Music likely represents a different demographic of listeners than Billboard, which only lists the bestselling records. The user-base of RYM could be loosely described as music enthusiasts, who enjoy listening to music from a critical standpoint, most likely listening to a larger variety of music than the average consumer. According to Similarweb.com, a web analytics service, 64% of RYM users are male and 65% are between the ages of 18-34 [38].

3.1.2 Data collection and variables

We obtained the historical billboard data set using a third-party library called "billboard.py", accessible from both GitHub and the Python Package Index (PyPI). The library utilizes web scraping techniques to fetch charts from the official Billboard website, allowing the retrieval of any Billboard chart on any defined date. [39]

The data retrieval process began with the earliest available Billboard 200 chart,

which was dated August 17, 1973. Weekly charts were systematically requested using a seven-day interval. Each weekly chart was consolidated into one table, and an additional column was introduced to denote the date attribute. The final data set spans from August 17, 1973, to December 31, 2023, and consists of 620,746 rows and 38,715 unique albums. The attributes contained in the billboard data are shown in Table 3.2.

Variable Name	Description	Type
artist	artist name	String
title	album title	String
isNew	has the album previously appeared on the chart	boolean
lastPos	last weeks position	int
peakPos	peak position	int
rank	current position	int
weeks	how many weeks	int
date	date of the chart	datetime

Table 3.2: Attributes in the Billboard dataset

To expand the data beyond having only the album titles and billboard rankings of a given date, data from Rate Your Music were obtained to enrich the data set with album metadata, such as the genre, release date, language, and average user ratings.

At the time of data collection, RYM is also lacking an official API, so another GitHub project called `rymscraper` was utilized to obtain data from the site [40]. To systematically obtain metadata for each album present in the historical Billboard data, the albums were grouped by artist and title, serving as their defining unique features. Data for each album was then requested from RYM through `rymscraper` and stored in json format.

Some artist names and album titles are formatted differently in the Billboard

data and Rate Your Music database, such as names that contain accented or special characters. To make sure that the fetched data could be later joined with the Billboard dataset, we preserved both the original titles as they appear on the Billboard chart, and the titles fetched from RYM.

Our method involved employing two different approaches to fetching data from Rate Your Music. Initially, we attempted to construct the URL directly using the artist name and album title. However, not all albums could be found with this approach. To make the data set more complete, we also utilized the site’s search function to get the correct url of a given album. While improving the overall completeness of the dataset, the latter approach occasionally yielded incorrect results, which will be discussed in Section 3.1.4. We excluded any soundtracks and compilation albums from this process due to facing problems collecting them from RYM. Compilations and soundtracks also generally do not correspond to single acts and genres, which is why they are not necessary for our analysis.

Variable Name	Description	Type
artistBillboard	artist name in Billboard	String
titleBillboard	album title in Billboard	String
artistRYM	artist name in RYM	String
titleRYM	album title in RYM	String
rating	average user rating	float
nRatings	number of user ratings	int
genres	list of genres	[String]
releaseDate	album release date	datetime
language	album language	String
descriptors	list of descriptors	[String]
releaseDate	album title	String

Table 3.3: Attributes in the RYM dataset

3.1.3 Web scraping ethics

Given that the data for this thesis were obtained using web scraping tools, it is essential to note that although it is a commonly used approach in data mining, web scraping can cause a significant load on targeted services when not used responsibly. Systematically scripted requests can overload a server designed for regular user interactions, which is why it's important to incorporate pauses between requests. Although this approach significantly extended the data retrieval time for this thesis, this is important to prevent possible server overload and ensure a responsible data collection process.

To assess the legitimacy of a research project that involves web scraping, previous authors have proposed the following set of questions that researchers should address [41]:

- Does the website's terms of use/service explicitly prohibit Web crawling or scraping?
- Does the website explicitly copyright its data?
- Does the project involve illegal or fraudulent use of the data?
- Can crawling and scraping potentially cause material damage to the website or the Web server that hosts it?
- Has the website sent the user a cease and desist letter, blocked the user's IP address, or closed access to data in some other way?
- Does the website's robots.txt protocol significantly limit or prevent Web-scraping activities?
- Can the data obtained from the website compromise individual privacy, research subjects' rights, or non-discrimination principles?
- Can the data obtained from the website reveal confidential information about organizations affiliated with the website?

- Can the project that requires the Web data potentially diminish the value of the service that the website provides?
 - Does the quality of the data obtained from the Web have the potential to lead to ill-informed decision making?
-

In case of this study, the data does not contain personal details or copyrighted material and as previously mentioned, additional steps were taken to ensure that no excessive stress was placed on the services. Concerns regarding web scraping typically revolve around parties attempting to profit monetarily from the data of other services, which is not the objective of this research.

3.1.4 Data cleanup and validation

As mentioned earlier, some artists and album titles can be formatted in various different ways. For example, names might contain accented characters, abbreviations such as *5 Seconds of Summer* to *5SOS*, and extended plays that might include the suffix (*EP*) at the end of the title. Due to these variations and the possibility that some albums were not found in RYM or incorrectly identified, it was necessary to develop a method to validate the results obtained from RYM.

The first step before comparing the strings was to standardize them to ASCII characters and convert all letters into lowercase. This approach already resulted in 84% of the strings matching. However, it still ignores any abbreviations and suffixes. We developed an algorithm to compare strings for possible abbreviations by iterating through both strings; the one present in the billboard data and the one in RYM data. Our algorithm also works in case of suffixes since a string lacking a suffix can be thought as an abbreviation of another string containing a suffix.

3.2 Analysis

Following the data collection and cleanup process, we began the data analysis process, aiming to address our research questions and gain insight on music trends and chart behavior. Our final dataset contains 620 746 rows in the historical Billboard data, spanning from August 17. 1963 to the end of 2023. The validated dataset gathered from Rate Your Music contains 26 695 rows, representing 69% of the unique albums present in the Billboard charts. The missing albums contain any soundtracks and compilations in addition to the albums we could not find using our methods.

In the following section, we will discuss the methods used to analyze and gain insight based on the data sets we gathered. All of the analysis was conducted in Jupyter notebooks using Python utilizing third party libraries, most notably NumPy, Pandas and SciPy.

3.2.1 Identifying unique albums and artists

Before combining the Billboard data with the RYM data, we wanted to make simple measures based only on the Billboard data set. This would allow us to gain a preliminary understanding of the music landscape and chart dynamics.

We began by using the `groupby()` function in Pandas, grouping the data by artist name and title, essentially getting rows of each unique album grouped as a result. We then aggregated the groups, taking the maximum and the minimum of the ‘date’ attribute, counting the appearances of each album, taking the maximum of the ‘weeks’ attribute and the minimum of the ‘rank’ attribute to get the peak position. As a result we would have a Pandas DataFrame containing each unique album, the date of it’s first and last appearance, the peak position it achieved, and the number of weeks it had spent on the list. To get similar statistics on artists, we grouped the albums by artist name, counting the number of albums for each artist and the dates of first and last chart appearance.

Next, we aimed to identify the most historically successful acts and albums from the data. To get the artists with the most Billboard top 1 albums, we filtered out each album with the peak position lower than one, then grouped the results by artist and counted the number of titles for each artist. To get the artists with the most weeks at number one, we filtered out all except the first album of each chart and counted the appearances of each artist. To identify the albums that have the most weeks on the chart, we utilized the ‘weeks’ attribute present in the original data, we also counted the number of appearances of each album, but this method was not consistent with the ‘weeks’ attribute, in some cases differing by a few weeks.

The artist names ‘Soundtrack’ and ‘Various Artists’ were excluded from the lists containing top artists, since they do not correspond to single acts but rather soundtracks and compilations. In addition to above, we looked at the top 10 most successful soundtracks and compilations which are listed in the appendix.

3.2.2 Distribution of chartlisted weeks among albums and artists

In addition to examining the most successful acts, we aimed to conduct quantitative measurements on the distribution of chart positions among albums and artists. Our goal was also to gain insights into how the distributions might have evolved over the years. To accomplish this, we began by dividing the dataset into two equal-length timeframes: from 1977 to 2000 and from 2000 to 2023. Next, we filtered out any albums that had appeared on the chart before the specified timeframe. We also filtered out artists ‘Soundtrack’ and ‘Various Artists’, since these do not correspond to single acts and would distort the measures. The timeframes were chosen due to our earlier results indicating that the amount of albums started growing significantly during the late 1990s and we wanted to capture how this change affected the distributions.

The factors we used to determine an artist's presence on the charts over time include the number of albums featured, the duration of each album's stay on the charts, and the cumulative weeks an artist appears on the charts. We wanted to visualize how these factors are distributed amongst albums and artists. To achieve this, we plotted each distribution in histograms. Computing the distribution of weeks per artist, we also took into account that multiple albums on the same chart only accounted for one week for the artist. In addition, we measured the mean, standard deviation and skewness of these distributions using SciPy, a Python library containing various statistical functions.

3.2.3 Genre analysis

To expand on the analysis based solely on the billboard dataset, we integrated the data gathered from Rate Your Music with the Billboard data. The amount of albums present in the RYM data set was 69 of the unique albums present in the Billboard charts. To analyze genre growth over time, our goal was to use the genre information available on RYM. However, we encountered instances where the genres provided were relatively obscure subgenres. To address this challenge, we utilized RYM's genre hierarchy, which categorizes genres into broader umbrella terms such as 'Pop' or 'Rock'. This hierarchy, obtained as a text document from GitHub [42], was parsed into a tree structure that was used to identify the root genres of each record. An album may still belong to multiple genres; for example, it could be classified as both Pop and Rock.

To visualize popular genres over time, we grouped the data by year and counted genre occurrences for each year. Because some of the data was unavailable on RYM. We counted the number of labeled data each year and divided the count of each genre by the number of labeled data that year. This approach allowed us to normalize the data, ensuring that the resulting visualization accurately represented the relative

popularity of genres over time.

3.2.4 Analyzing RYM ratings

Trying to understand music genre growth has been a common topic in music trend analysis, but one unique feature present in the RYM dataset is the mean rating based on the user ratings. For each album, users can give a rating on a scale from one to five stars. The mean rating can be used as a feature to understand how these music enthusiasts view the quality of past and contemporary popular music.

To gain understanding of how albums are typically placed on the scale from 1 to 5, we visualized the distribution of all the ratings and calculated the mean value. Next, we grouped the data by genre calculating the mean rating of each genre and visualized the results in a bar chart. Additionally, we took the mean rating of each genre and calculated the corresponding quantile of that value by dividing the amount of ratings below the mean by the total number of ratings. In addition we grouped all unique albums by genre and visualized the distribution of ratings across the most popular genres. To reinforce our results regarding the difference among various genres, we utilized the Kruskal-Wallis test to determine whether the distributions are significantly statistically different. We chose to do a pairwise comparison among the 9 most popular genres to determine the significance between all pairs.

Similarly to how genres were visualized, we grouped the data by year, and computed the mean rating for each group. Additionally, we also computed the average number of ratings, as that is essential for understanding potential biases associated with past records. To visualize the correlation between the rating evolution and trend growth, we also computed mean ratings over time for the most historically popular genres.

4 Results and Findings

4.1 Results of Billboard analysis

4.1.1 Yearly unique albums

In total, the historical dataset from billboard contains 38 715 unique albums. Over the course of the 60 years in our data, this would mean around 645 new albums would appear on the chart each year per average. Figure 4.1 visualizes the amount of unique records appearing on the chart each year, and the amount of new records that have previously not appeared on the chart. From the figure, it's apparent that the amount of both new and total yearly albums grew rapidly from around 1990 to the year 2014, but has since declined. In Figure 4.2, we visualized the proportions of records that have previously appeared on the charts over two years ago, and over ten years ago. The figure exhibits that the amount of older hit records has suddenly started growing since around 2008.

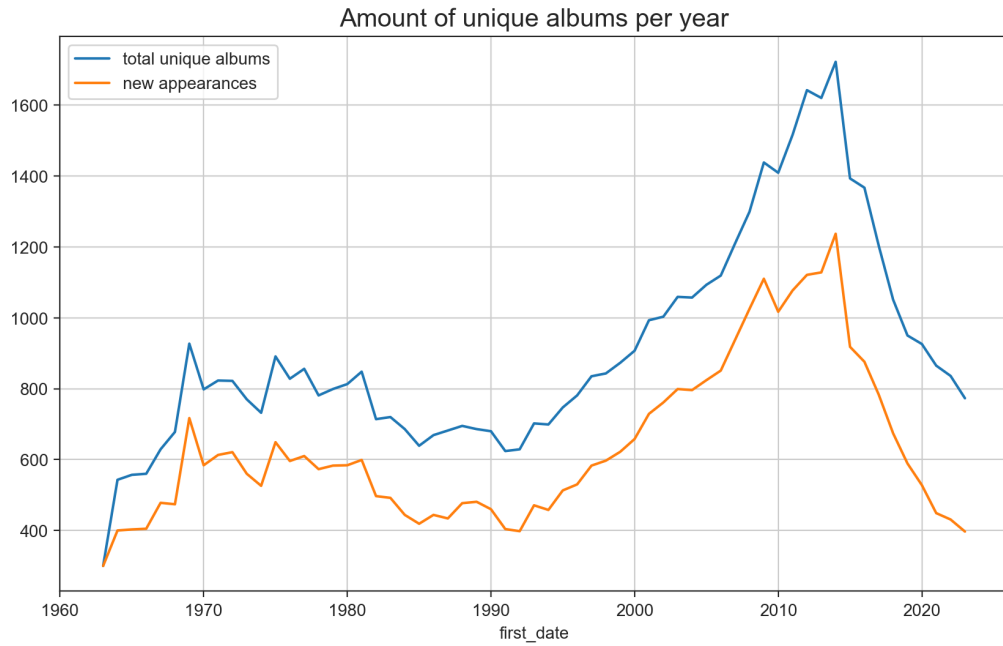


Figure 4.1: Visualising the number of unique albums appearing on the Billboard 200 each year. The blue line represents the total number of unique albums each year, while the orange line represents unique albums that have not appeared on the chart prior to that year.

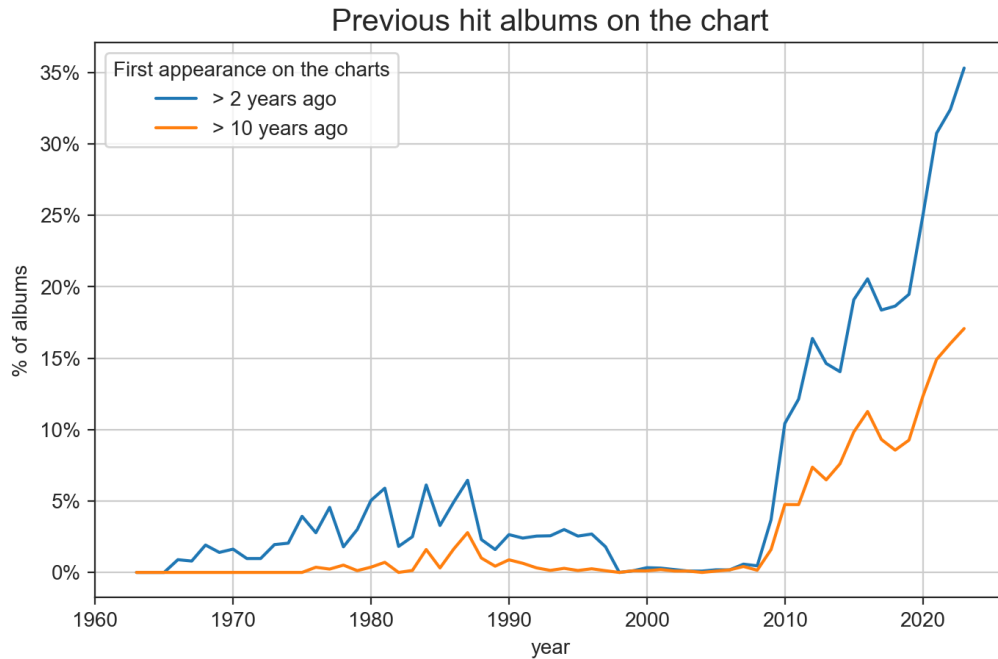


Figure 4.2: The yearly percentage of previous Billboard hit records appearing on the chart. The records that might have been released earlier but have never appeared on the chart are ignored.

4.1.2 Most Successful Acts

In Table 4.1, we present the top 10 artists with the most Billboard No. 1 albums in our data set. In particular, The Beatles led the ranking with 19 unique albums, highlighting their significant impact on the music industry. Hip hop and rock emerge as the most common genres in the top 10. The Beatles and The Rolling Stones are the two British acts in the top 10, with the rest of acts being from North America.

Artist (solo/group)	No. Albums	Years active	Genre(s)	Country
The Beatles (G)	19	1960-1970	Rock	UK
Taylor Swift (S)	13	2004-	Pop, Country	USA
JAY-Z (S)	11	1986-	Hip hop	USA
Bruce Springsteen (S)	11	1964-	Rock	USA
Drake (S)	11	2001-	Hip hop	Canada
Barbra Streisand (S)	10	1960-	Broadway	USA
Eminem (S)	9	1988-	Hip hop	USA
Madonna (S)	9	1979-	Pop	USA
Kenny Chesney (S)	9	1993-	Country	USA
The Rolling Stones (G)	9	1962-	Rock	UK

Table 4.1: Artists with the most Billboard top 1 albums since 17.8.1963

Looking at the acts with the most weeks on Billboard No. 1, The Beatles again hold the top spot in Table 4.2, followed by Taylor Swift. These results highlight the commercial success of Swift in the 21st century, surpassing many well-established acts in terms of Billboard chart presence. This emphasizes her significant influence and popularity in the contemporary music landscape.

Transitioning from artists to single records, Table 4.3 shows that Pink Floyd's *The Dark Side of The Moon* (1973) holds the record for the most weeks spent on the chart overall, totaling 987 weeks (nearly 19 years). On average, an album appears on the chart for a total of 16 weeks. Additionally, we looked at the time span between the first and last appearances on the chart. The average was 42 weeks, while the longest duration was over 56 years with The Beatles' *Revolver* (1966) which made its first appearance on the chart in March 1966, and last on January 2023.

Artist (solo/group)	Weeks	Years active	Genre(s)	Country
The Beatles (G)	132	1960–1970	Rock	UK
Taylor Swift (S)	67	2004–	Pop, Country	USA
Garth Brooks (S)	52	1985–	Country	USA
Michael Jackson (S)	51	1964–2009	Pop	USA
Whitney Houston (S)	46	1977–2012	R&B, Pop	USA
Adele (S)	40	2006–	Pop	UK
Elton John (S)	39	1962–	Rock	UK
The Rolling Stones (G)	38	1962–	Rock	UK
Fleetwood Mac (G)	38	1967–	Rock	UK
The Monkees (G)	37	1966–2021	Rock	USA

Table 4.2: Artists with the most cumulative weeks at No. 1 since 1963

Artist	Title	First appearance	Weeks
Pink Floyd	The Dark Side Of The Moon	1973-03-17	987
Bob Marley And The Wailers	Legend: The Best Of Bob Marley And The Wailers	2023-11-11	814
Journey	Journey’s Greatest Hits	1988-12-03	786
Metallica	Metallica	1991-08-31	735
Creedence Clearwater Revival	Chronicle: The 20 Greatest Hits	1976-03-06	673
Eminem	Curtain Call: The Hits	2005-12-24	663
Bruno Mars & Doo-Wops	Hooligans	2010-10-23	655
Guns N’ Roses	Greatest Hits	2004-04-10	655
Nirvana	Nevermind	1991-10-12	649
Michael Jackson	Thriller	1982-12-25	605

Table 4.3: Top Albums by Weeks on Chart

4.1.3 Distribution of chart presence

To measure how chart presence is distributed among different artists and albums, we visualized the distributions of albums per artist, chart weeks per album and chart

weeks per artist. The data was split into two parts: charts from 1977 to 2000, and 2000 to 2023. Any albums that appeared on the charts before the given timeframe were also removed from the respective subset. The data previous to 2000 contains 10 625 unique albums, while the later subset contains 17 800. These numbers again tell us how the number of records has increased since the late 1990s.

From Figure 4.3, it's apparent that the distribution of unique albums per artist is largely right-skewed, indicating that fewer artists have their albums appear on the chart consistently. The skewness is also noted in Table 4.4, which tells us that the distribution among charts post 2000 is more skewed than the charts before.

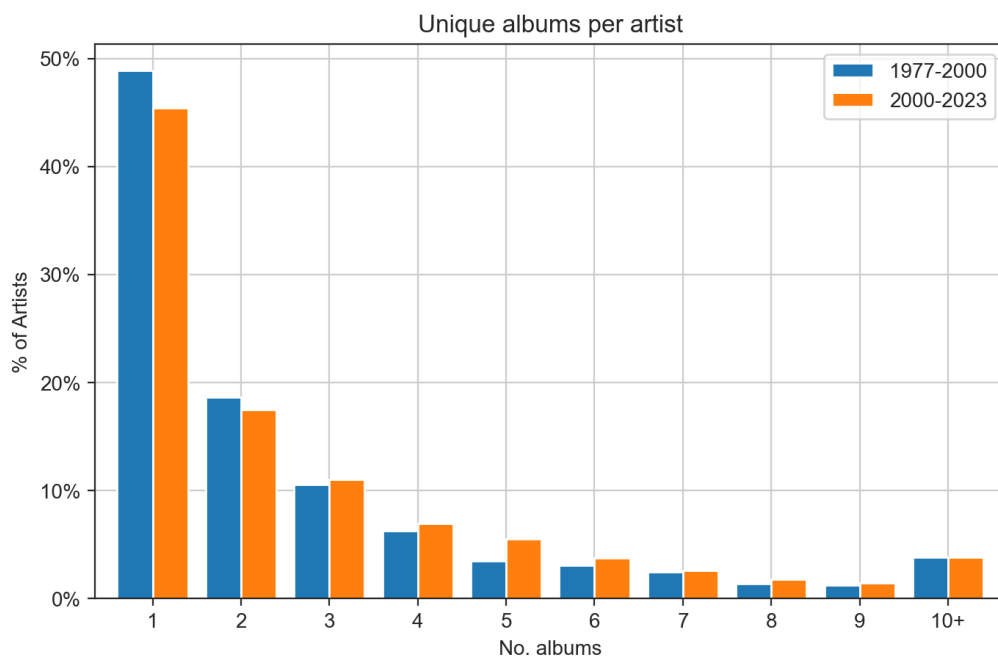


Figure 4.3: Distribution of charting albums per artist in the Billboard 200 chart

Figure 4.4 exhibits how chart weeks are distributed per albums. Again it's instantly noticeable how skewed the distribution seems to be. Judging by this figure, it's also clear how the distribution has changed between these two timeframes present in the figure. The rising amount of unique albums appearing on the chart (Figure 4.1) seems to correlate with a larger share of albums appearing on the chart for less than one month.

Looking at the distribution of chart listed weeks per artist in Figure 4.5, it's again apparent that a small percentage of the artists account for the most weeks on the chart. The distribution also has become more skewed since the 2000's, with over 60% of artists appearing on the list for less than a year.

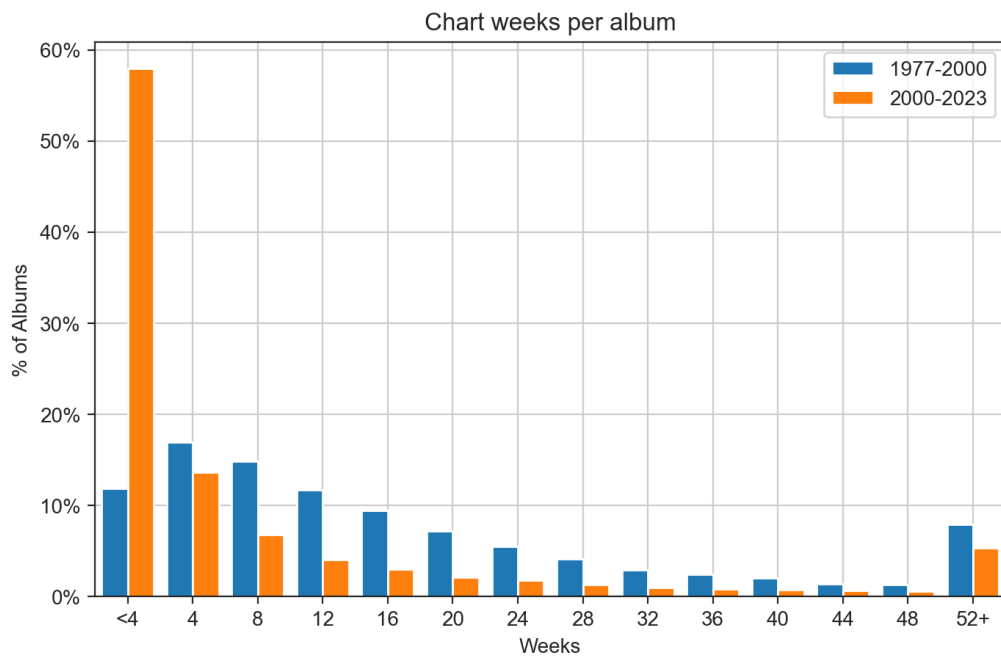


Figure 4.4: Distribution of chart weeks per album in the Billboard 200 chart

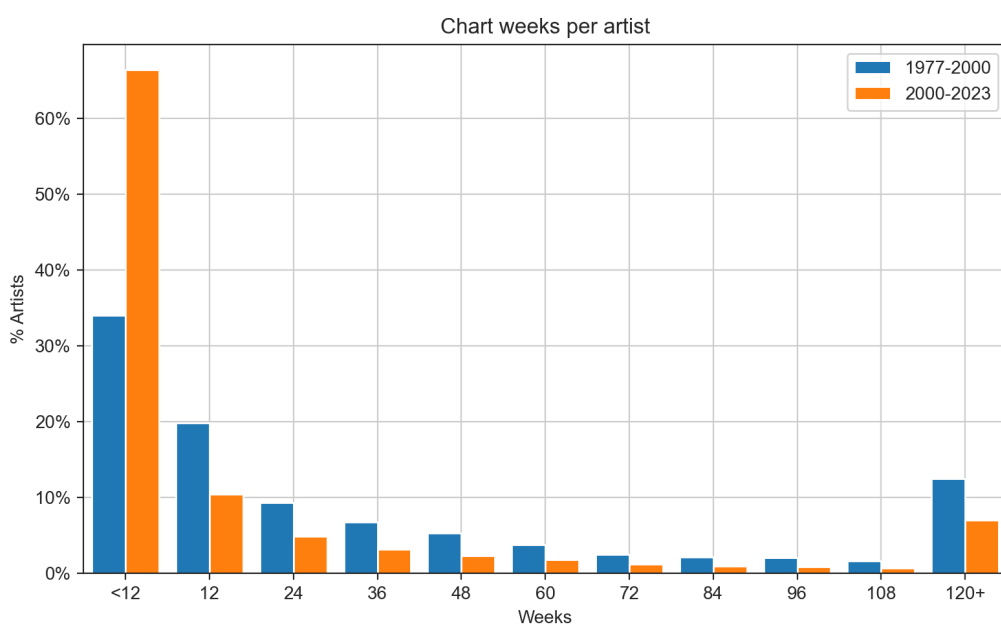


Figure 4.5: Distribution of chart weeks per artist in the Billboard 200 chart

Timeframe	1977-2000			2000-2023		
	Mean	Stdev	Skewness	Mean	Stdev	Skewness
No. Albums	2.71	2.90	2.98	2.92	3.27	5.98
Weeks per album	20.55	23.29	3.36	12.16	32.55	8.18
Weeks per artist	51.74	80.33	3.20	30.96	79.51	5.25

Table 4.4: Descriptive statistics of chart presence per album and artist

4.2 Genre proportions of charting albums over time

In this section we will discuss the genre proportions present in the Billboard data set. Genres for each album were fetched from Rate Your Music and we will use the categorization provided by the site. Not all albums could be collected from the site so it's important to keep in mind that the data might be slightly biased towards more prominent records from the past.

4.2.1 Overall genre distribution

Looking at the overall genre distribution of our data set in Figure 4.6, Pop is the largest genre overall, accounting for 22.7% of the data. Rock follows as the second largest with a 19.0% share, and R&B ranks third with 12.7%. One important note is that in our data, which relies on the RYM genre classification, albums can be classified as belonging to more than one genre.

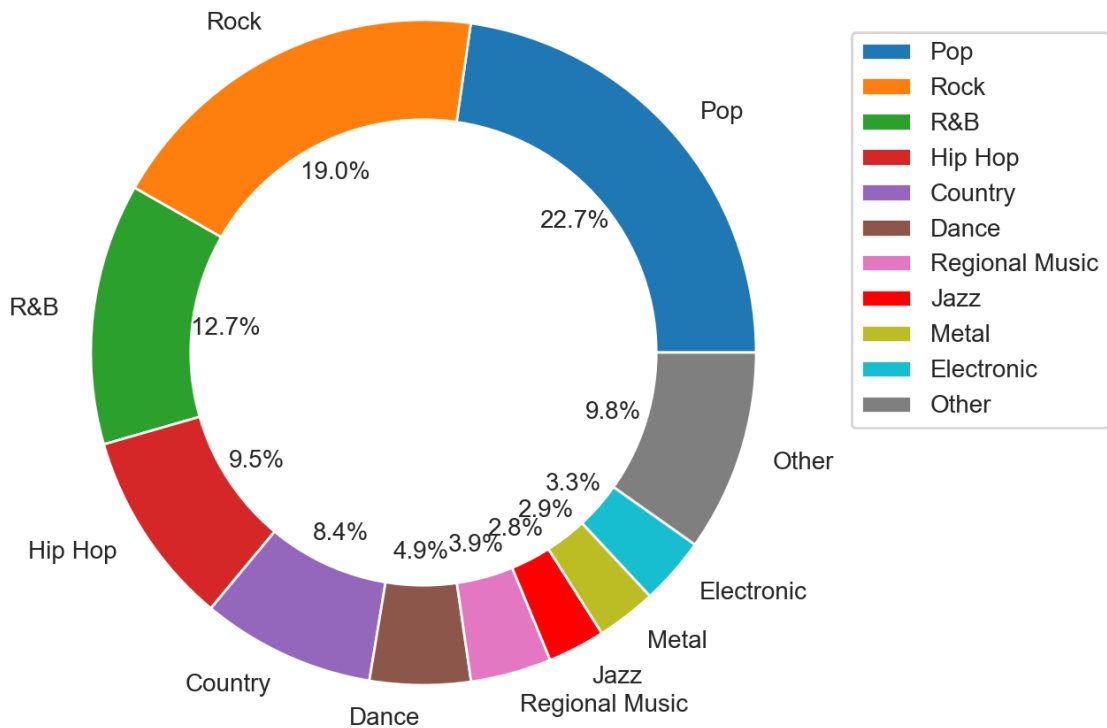


Figure 4.6: Overall genre distribution among the unique albums that have appeared on the Billboard 200

4.2.2 Genre distribution over time

To gain more insight into the most prevalent genres during each timeframe, we visualized the distribution of genres over time in Figure 4.7. From this figure, we observe that the largest genres pop and rock, dominated the charts up until 2018, when hip hop became the most prevalent genre. The relative popularity of both pop and rock also saw a large drop during the 90s, where hip hop and country first saw a sudden rise in popularity.

Other popular genres include R&B, country, dance, regional music, jazz, metal, and electronic. of which R&B has been the most popular up until the 90s, where hip

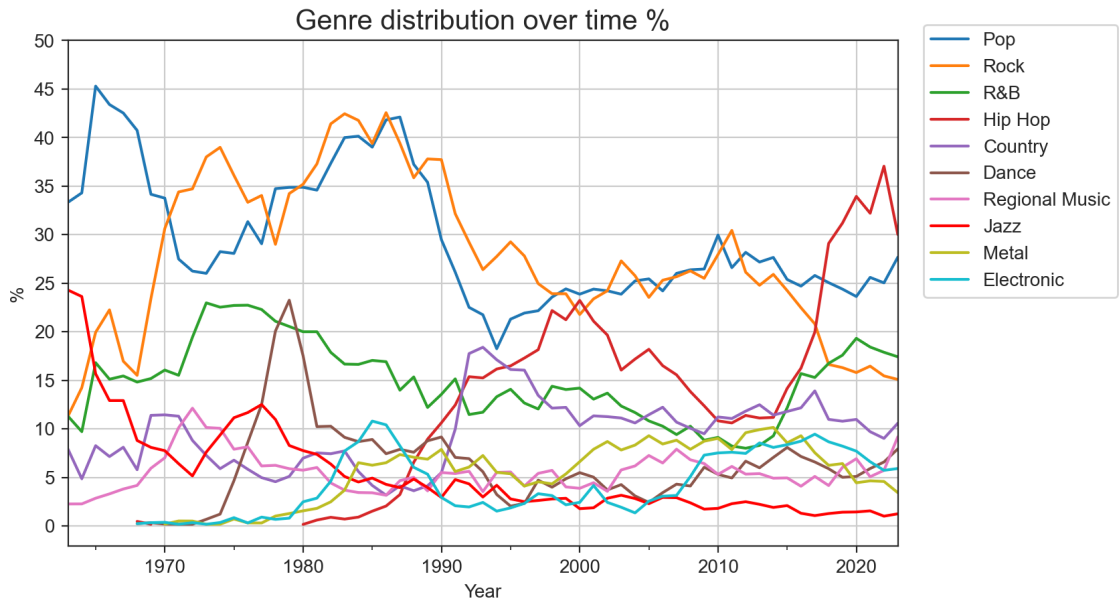


Figure 4.7: Genre growth over time. The figure is based on unique albums each year not depending on the number of weeks

pop first started gaining traction and later became the most popular genre overall. The remaining top 10 genres have maintained a popularity of around 5-10%.

4.3 User ratings

Looking at the user ratings present in the RYM dataset, we initially computed the mean over all unique albums, which was 3.16 stars. We also plotted the distribution of ratings across all the albums, which is shown in Figure 4.8. The figure shows that the distribution is slightly skewed to the left. We also calculated the standard deviation and skewness, which were 0.5 and -0.84, respectively. The negative skewness and a mean of 3.12 indicate that the top 200 albums receive slightly more positive values than the median rating which would be 2.75 stars on the scale from 0.5 to 5 stars.

To address the question about how people view contemporary music compared to the past, we visualized the mean rating for each year. Figure 4.9 illustrates

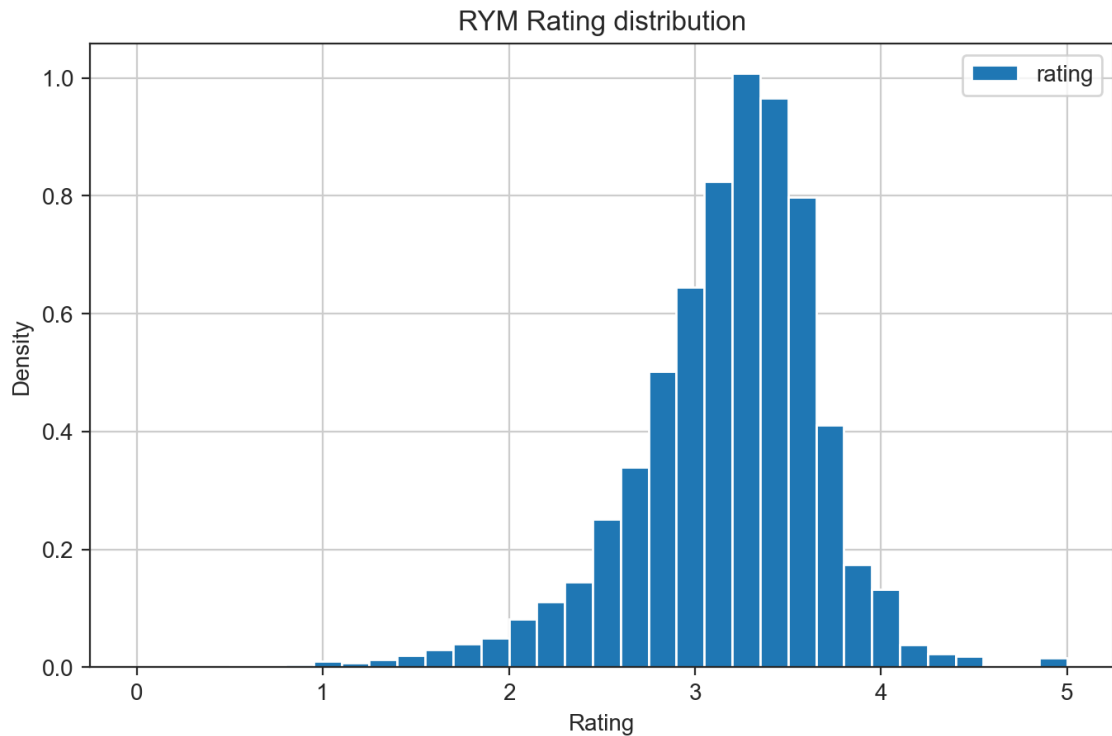


Figure 4.8: Distribution of RYM Ratings

that during the late 1990s, there was a relatively significant drop in the average rating, decreasing from around 3.3 to 3.1 and below. In addition to mean, we also visualized the 95th and 5th quantiles, which tells us that the top 5% highest rated albums remain around the same rating, and the drop in ratings is most drastic in the lowest 5%. This suggests that music enthusiasts perceive past records as higher quality than those of the 21st century.

Considering ratings per each genre, we visualized the mean rating for the top 10 most popular genres in our dataset. The results in Figure 4.10 reveal that among the most popular genres, rock received the highest mean rating, with a mean rating of 3.39 stars, while hip hop received the lowest rating, at 2.91 stars. Interestingly, the "other" category, which includes all records outside of the biggest genres, is placed second with a mean rating of 3.38 stars. Additionally, the mean ratings are listed in Table 4.5, with the corresponding quantiles. The results show that for example the

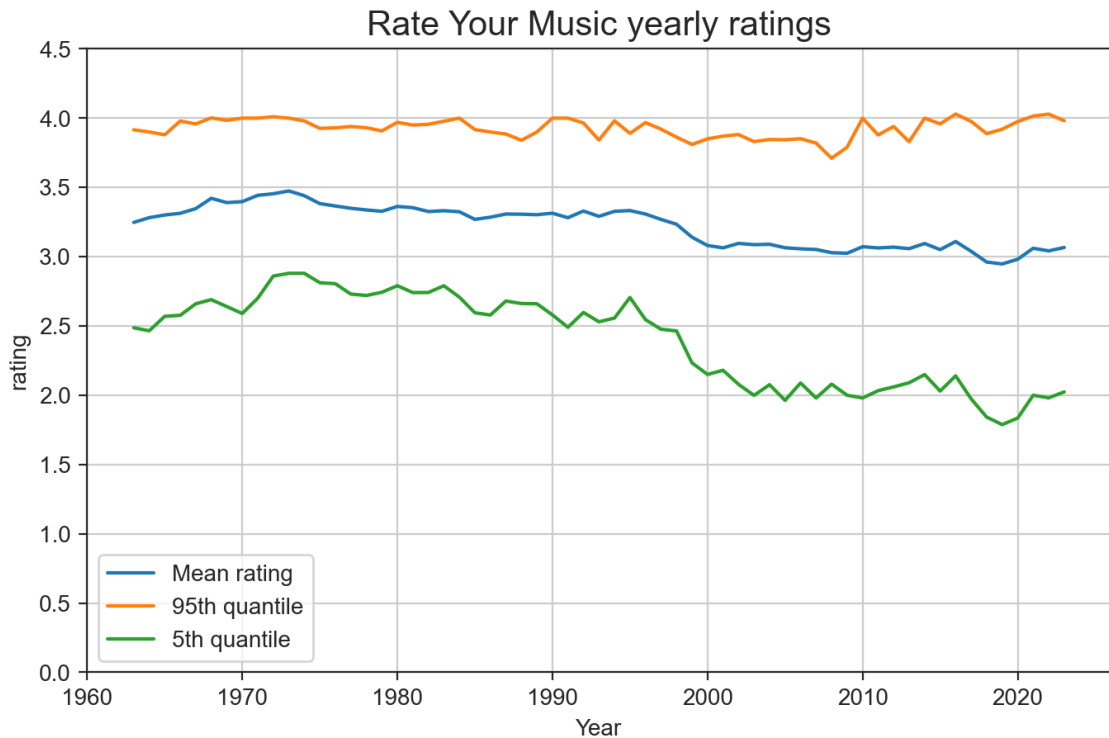


Figure 4.9: Yearly mean rating and rating quantiles of unique Billboard 200 albums. Each album represents a single data point not depending on how many weeks it appeared on the chart

mean rating of rock (3.39) corresponds to the 65th quantile of all ratings, meaning that 65% of ratings are scored below the mean rating of rock. On the other hand, only 25% of albums are scored below 2.91 stars, the mean of Hip hop albums.

To further examine the differences in ratings between genres, we visualized the distribution of ratings across the 9 most popular genres in Figure 4.11. In addition to the visualization, we conducted pairwise Kruskal-Wallis tests for the most popular genres. Table 4.6 shows the resulting p-values from each comparison, showing that all of the distributions excluding Regional Music and Country, and Metal and R/B are significantly different statistically.

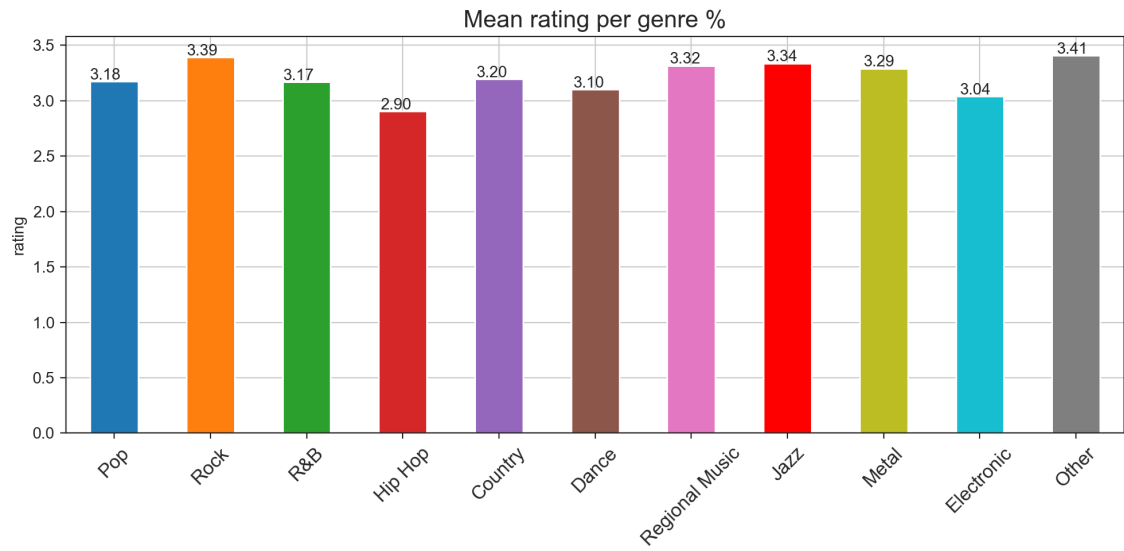


Figure 4.10: Mean rating for 10 most popular genres

Genre	Mean rating	Quantile
Pop	3.17	0.43
Rock	3.39	0.65
R&B	3.16	0.43
Hip Hop	2.91	0.25
Country	3.19	0.45
Dance	3.10	0.38
Regional Music	3.27	0.55
Jazz	3.33	0.59
Metal	3.29	0.56
Electronic	3.03	0.33
Other	3.38	0.64

Table 4.5: Mean ratings of the most popular genres and the corresponding quantiles

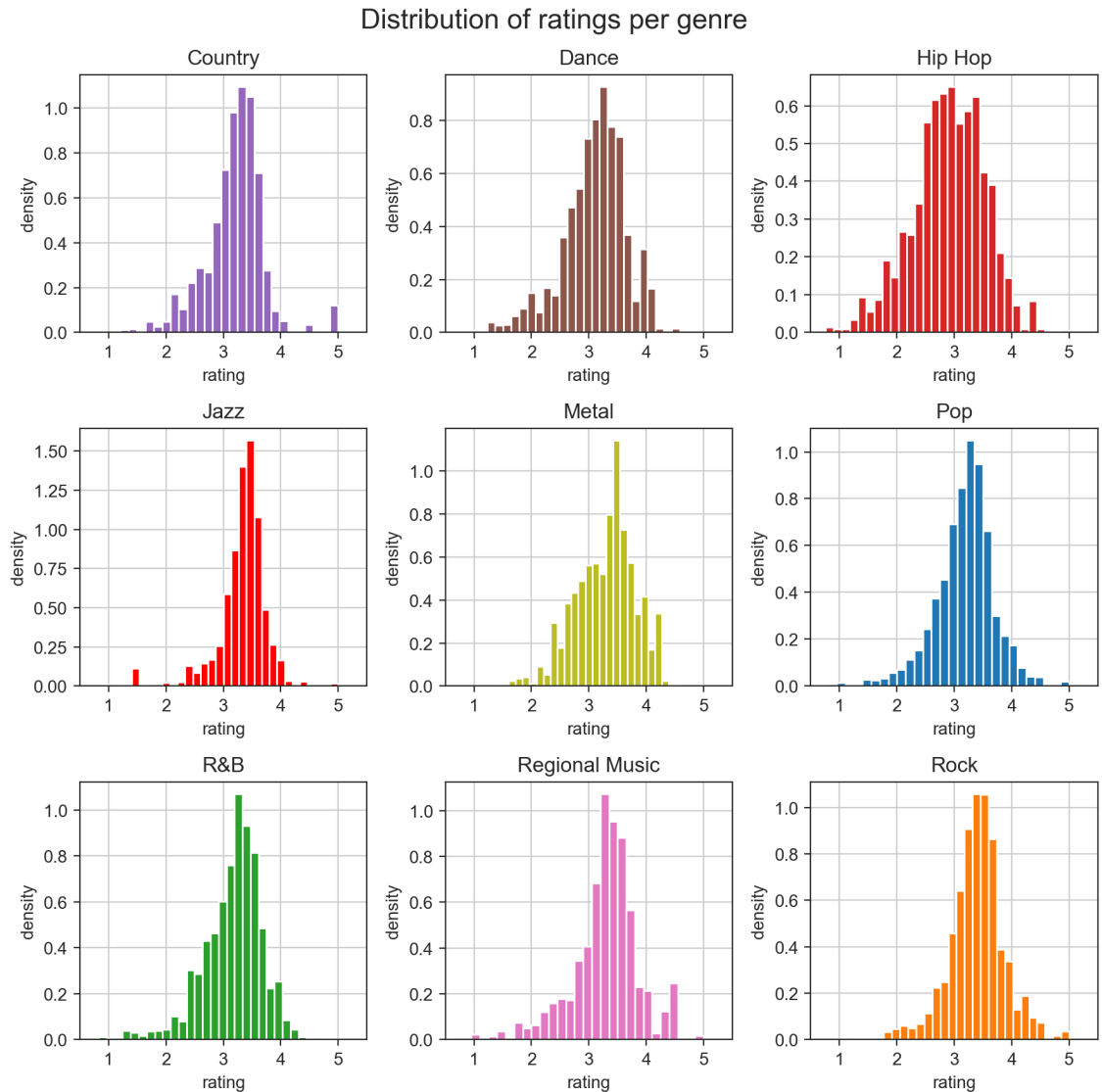


Figure 4.11: Distribution of RYM Ratings for each of the top genres

We also calculated the mean ratings per genre for each decade, and visualized it in Figure 4.12. This figure shows a similar pattern to the overall rating trend, with each average dropping significantly from the 1990s to the 21st century. Metal seems to have had the highest mean rating of all time during the 1970s, but as seen in Figure 4.7, the share of metal was quite small during that time period. Notably, hip hop, which became the most popular genre from the late 2010s to the present, has received the lowest average rating during this time.

Genre	Country	Dance	Hip Hop	Jazz	Metal	Pop	R&B	Regional Music	Rock
Country	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.62	0.00
Dance	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Hip Hop	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
Jazz	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
Metal	0.00	0.00	0.00	0.00	1.00	0.00	0.18	0.00	0.00
Pop	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
R&B	0.00	0.00	0.00	0.00	0.18	0.00	1.00	0.00	0.00
Regional Music	0.62	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Rock	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Table 4.6: Pairwise Kruskal-Wallis test p-values for comparing the rating distributions across different genres.

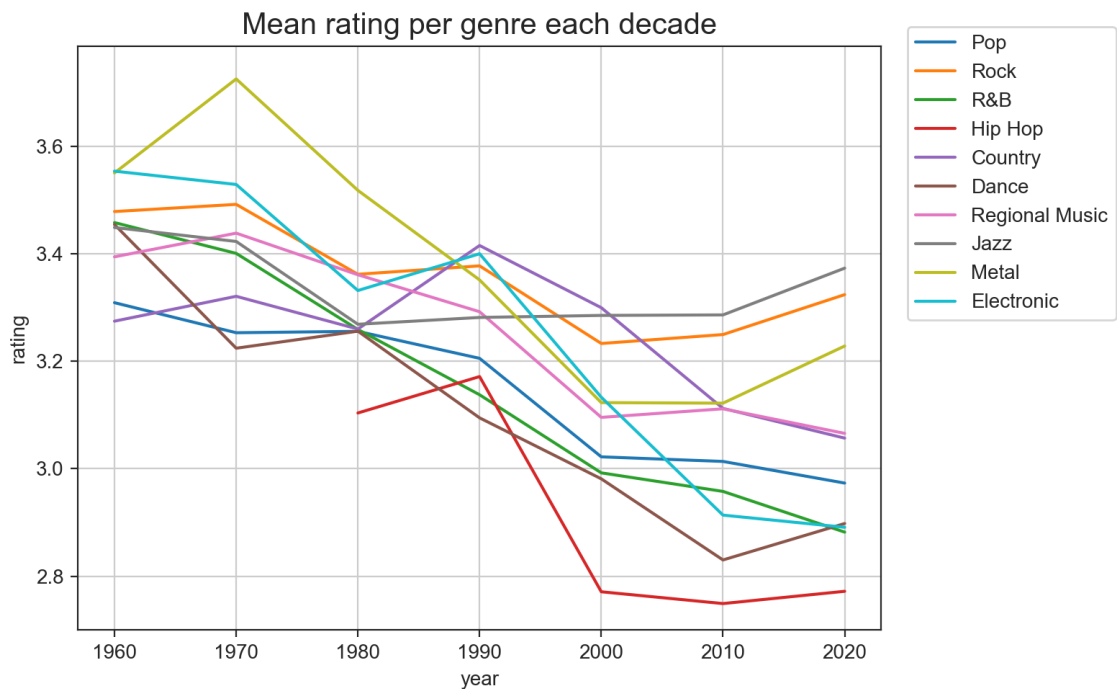


Figure 4.12: Mean rating by genre for each decade

5 Discussion

In this thesis, we collected a historical dataset containing charts of the top 200 best-selling albums in the US. We also collected album-level metadata from Rate Your Music, containing features such as genre, language and rating. To explore patterns and trends present in the data, we used various statistical methods and visualizations, focusing on the interesting aspects of the datasets. In this section we will compare our results to previous studies and discuss the potential implications. We will consider the relevance of our results regarding previous literature and discuss the potential contributions of our research for both future research and practical applications. Additionally, we'll cover any limitations of our study and suggest possible improvements for future research.

5.1 The variety of yearly unique album appearances on the Billboard 200 chart

In section 4.1.1, we examined the number of unique albums appearing on the charts yearly. It's apparent that the number of yearly unique albums began rising in the 1990's, coinciding with the popularization of CDs. Figures 5.1 and 5.2 show the revenue and volume of recorded music in the US [43]. Comparing our results to these figures, it's possible that the growth in album diversity present in the Billboard 200 charts was a result of increased sales volumes and revenue brought in by the CD

format. While the shift to digital formats impacted revenue, sales volumes grew rapidly and the amount of unique albums kept rising. Interestingly, as paid streaming subscriptions started gaining mainstream popularity around 2014, increasing the recorded music revenue, sales volumes decreased and so did the amount of unique albums in the charts, as seen in Figure 4.1.

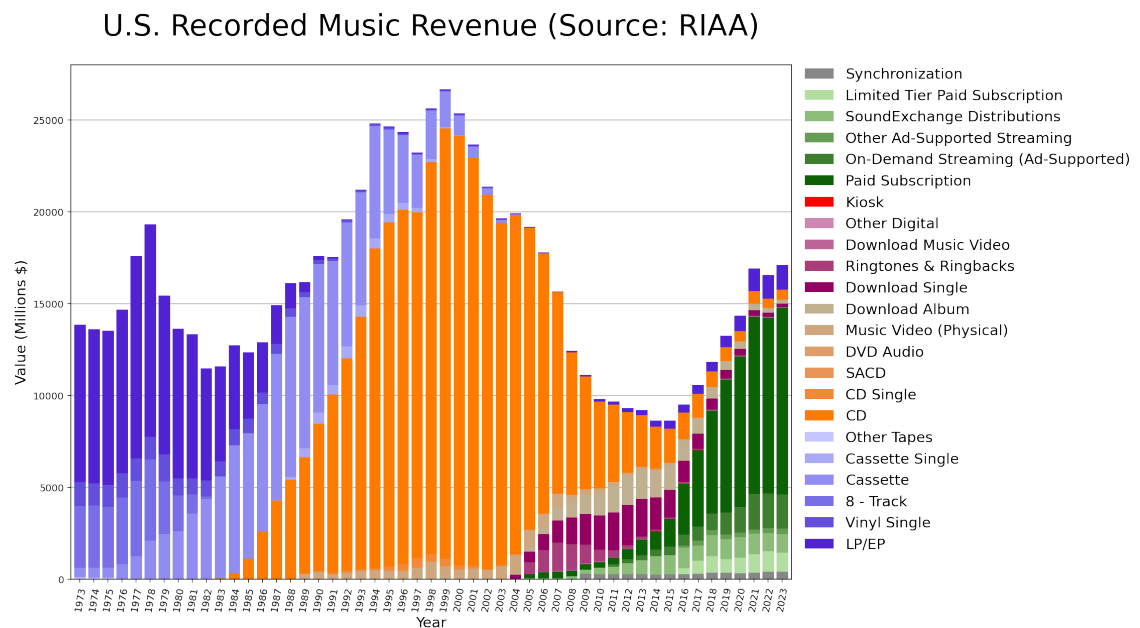


Figure 5.1: Yearly revenue of the US recording industry (Adjusted for inflation) based on data from the RIAA Music Database [43]

5.1.1 The return of previous hit records

In section 4.1.1, we also examined the share of older hit records present in the charts. Interestingly, the percentage of older records in the charts has seen a sudden rise since 2008, now accounting for a large part of album sales. The growth lines up with the increase in the recorded music revenue in the US, and we think the majority of it is a direct result of the album replacement cycle and the constant revenue brought in by digital streams.

Before the current era of digital media, if a person wanted to listen to a record that was charting two years ago, there was a chance they would already own it, or

U.S. Recorded Music Sales Volumes (Source: RIAA)

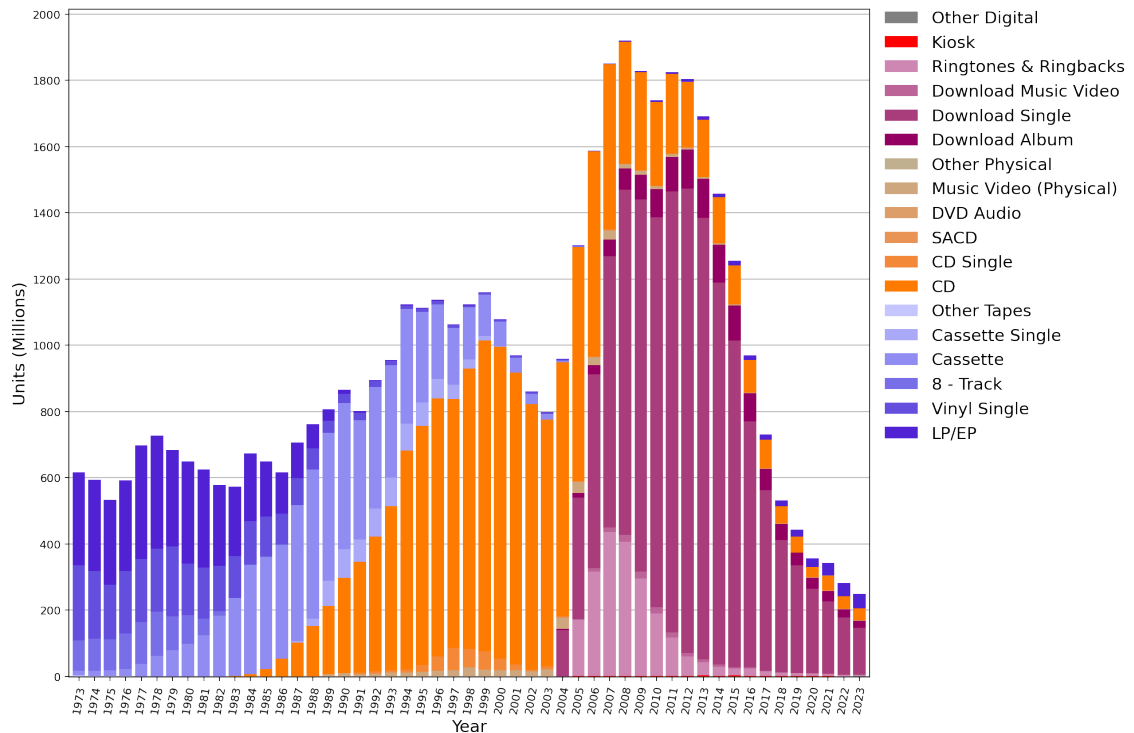


Figure 5.2: Record sales volumes in the US based on data from the RIAA Music Database [43]

they could buy it second-hand. Neither of these options bring in sales or revenue for the industry. However, with digital streaming, the listener creates the same amount of revenue per stream listening to 50-year-old records as they would listening to brand new releases.

In his article *Is Old Music Killing New Music* (2022) [44], Ted Gioia suggests that the increase of older music present in the charts is more than a statistical anomaly resulting from increased digital sales. Gioia speculates that major record labels are losing interest in new music due to the rights to old records being just as profitable, creating revenue seemingly from nothing. He also writes that record labels are actually rushing to buy these rights to old records, instead of investing in new artists. The growing interest in older music can also be seen translated to the movie industry with the success of recent documentaries such as *Bohemian*

Rhapsody (2018) and *The Beatles: Get Back* (2021), which in turn can lead to older releases rising back to the charts as seen with The Beatles' *Revolver* (1966) making a comeback on the charts after over 52 years since release.

5.2 The most successful acts and the distribution of chart presence among artists

In section 4.1.2 we looked at the acts and albums that have been the most prominent on the chart based on different measures. The most prominent acts are all from English-speaking countries, which is to be expected considering Billboard charts are based off of sales in the US. We also saw rock as the most prominent genre in these listings with a large portion of the all-time prominent acts being rock bands from the 1970s and 1980s. However, looking at the artists with the most top 1 records, there are more recent artists such as Drake and Taylor Swift already contesting many of the well-established acts of the 20th century.

To understand what makes these contemporary artists able to compete with the older acts that have had decades of more time to establish their image, we looked at how the chart presence is distributed among artists, dividing the data into two, before and after the 21st century. Our results showed that in all three categories we measured, since the 2000's, the distribution of chart presence has become more skewed towards favoring a tiny portion of artists. The increased skewness in the distributions in the 21st century could also be a result of more unique albums appearing on the list overall, which would mean that the lower portion of the chart has become more volatile while the top portion is still controlled by a few of the most popular acts.

We intentionally took a similar approach to measuring these distributions as Strobl et Al. [45], who measured the chart dynamics in the UK during the years

1993-2000. Our goal was to repeat their process on US data and see how the results would compare. The authors of the paper found out significant skewness among the distributions, concluding that it might serve as indirect evidence of the bandwagoning effect, where the initial success of an artist further snowballs their popularity. The authors also point out that from the perspective of recording labels, investing in the success of already popular acts is less risky and more profitable than investing in new artists.

Comparing our results to the results in UK during 1993-2000, all of the distributions appear more skewed in the US charts both before and after the 2000s. We think that the increased skewness since the 2000's might be a sign that during the age of social media and music streaming, the bandwagon effect might be further amplified, due to social media and the recommender algorithms of streaming services functioning as an echo chamber to boost the popularity of already prominent acts.

5.3 Most prominent genres accross the history and the value of genre as a music classification tool

As previously discussed, there are various ways to classify music into different genres, and there is not a single taxonomy. It's important to note that we chose to use the genre taxonomy present in Rate Your Music and all of the analysis is based off of that. As RYM is a community driven platform, the taxonomy is a result of collective effort made by the site's users, rather than something made up by a single organization. Another important note is that while categorizing music into the broad umbrella terms we used in this study, the classification becomes inherently reductive, ignoring the existence of various subgenres present under these umbrella categories or root genres. For example, in Figure 4.7 the popularity of R&B seems to have peaked in the 70s, while the general understanding is that contemporary

R&B as a genre was born during the late 80s and 90s and has influences from other genres of that time, most notably hip hop. This is due to the fact that the RYM genre taxonomy categorizes genres such as soul and funk under the R&B category.

Earlier research on genre trend analysis has focused mainly on the single charts [21]–[23]. Whereas the Billboard 200 represents album sales, the Billboard Hot 100 single chart represent the radio plays and streams. Comparing our results to the results by The DataFace [22], it’s apparent that the trends are more exaggerated in the singles list. In the Billboard Hot 100, for example the popularity of rock peaks at 60.9% of chart spots during the 1983, but falls under 10% during the 2000s. The same trend is present in our data, but the changes are less dramatic. When comparing the results, it’s important to keep the differing genre taxonomies in mind, for example the study by The DataFace makes the distinction between Soul and R&B.

In general, such analysis of genres ignores the stylistic evolution of the genres themselves. Contemporary hip hop for example, is in many ways different from hip hop of the 90s. Being the most popular genre today, hip hop has taken a more mainstream form, taking influences from other currently popular genres such as electronic music and pop. For example, contemporary hip hop has put more emphasis on the hook or chorus of the song, while simultaneously sacrificing lyrical complexity due to less emphasis on the verses [46].

To develop more profound understanding of stylistic evolution of music, the genre-level analysis could be combined with music information retrieval methods to evaluate the changes within the genres. Mauch et Al. [23] combined user tags from Last FM with MIR methods to detect these changes in the Hot 100 from 1960 to 2010. The researchers measured various musical ‘topics’ related to harmony and timbre, such as chord structures, percussion elements and brightness or energy. They found that for example, the topic ‘no chords’ could be associated with the

rising amount of Hip hop singles in the 1990s.

However, defining such features regarding albums is a more complex task than singles. MIR methods could be applied for each song, but the relationship between songs would also need to be considered to represent the album as a whole. This would also require a large amount of data, since we would be processing entire albums instead of single songs.

5.4 User opinions in Rate Your Music and the perceived quality of music accross different decades and genres

The results on the mean rating over time show a relatively drastic decrease in the ratings during the late 1990s. This indicates that the online users in general view older music as being ‘better’ or at least more enjoyable. The same trend is carried over to each genre, with almost all of the most popular genres experiencing at least some form of decrease in the ratings from the 1990s to the 2000s. Country is the only genre that maintained it’s mean rating, with hip hop experiencing the most significant decrease.

The variance in overall ratings per genre is also significant. Rock has the highest mean rating, placing the average in the 65th quantile of all albums. The other genres have around 3.0 to 3.3 mean ratings, the only exception being Hip hop, averaging a 2.91 rating. Although the differences in the mean ratings might seem small, comparing them to the overall distribution of ratings reveals that the mean rating of hip hop is placed in the 25th quantile of the data, implying that on average, hip hop albums are less enjoyable than 3/4 of the charting albums.

While going over the background for this thesis, we covered the value of nostalgia in music, but what’s interesting is that considering most of the users in RYM are

between the ages 18-34, the high rating of older music contradicts the theory that listeners enjoy music popular during their young adulthood. It's possible that users around this age can experience nostalgia for those records from for example their parents playing older records during their childhood, but it's not a likely reason for the difference in ratings.

We have already covered over how various technological changes have altered the music industry, regarding the production, distribution and consumption of music. The decrease in user ratings seems to line up with the increasing popularity of compact disks, the increase in album variety in the charts and the increased music revenues brought in by the CD. From the data we have, it's impossible to tell whether these factors are the cause for the decrease in ratings, but there definitely seems to be a correlation between these factors. However the increase in the number of albums could suggest that record labels were more focused on quantity over quality, leading to many records being less thoughtful and thus less enjoyable from the perspective of music enthusiasts.

Considering that rock has the highest average rating and the decrease in ratings happened after the number of charting rock records decreased, it's possible that the decrease in album scores is a sign of 'rockism' present in Rate Your Music's userbase. Rockism is a term used to describe the belief that rock music is inherently more authentic and artful than other genres like pop and hip hop [47]. However, rock experienced a similar downfall in ratings since the 1990s than other genres, which could be an indication that rockism, at least regarding popular music, applies only to the classic rock records made in the 1970s to 1980s. This is further supported by the fact that many of the longest charting records are rock albums made during that time, and all time most prominent acts in the Billboard 200 are rock acts that reached popularity during 1970s and 1980s.

5.5 Theoretical and practical contributions

We have explored various topics related to the evolving trends of the music industry. Our results regarding genre popularity across the last few decades provides understanding of the evolving nature of the popular music landscape, but can also function as a theoretical background for a more in-depth analysis of genres. Several studies have already conducted similar research on genre growth over time, but being the most recent, our study shows the recent mainstream success of hip hop. Most of the existing literature focuses on single charts, while our study provides a source for comparison between genre trends in single vs album releases.

We also explored the changes brought by technological development in the music recording and distribution field, which can help consumers and recording labels understand how different music distribution technologies affect listening habits and also the way music is produced and marketed. The analysis of chart presence distribution shows how a small portion of artists are occupying a large part of the charts.

The analysis on online user based album ratings provides a general understanding of popular music from the perspective of music enthusiasts. We've explored how the ratings are distributed across music from past decades and various different genres. These results could also function as theoretical insight for future work on how different types of listeners enjoy music from various perspectives, for example what features make contemporary popular music appealing for the general audience, but less enjoyable for the enthusiasts.

5.6 Limitations and future work

Due to our data sources, this study was limited to only the best-selling albums which of course do not represent the entire recording industry. For example, calling out

Rate Your Music users for being biased towards rock albums in general would be incorrect, and even though the ratings for charting hip hop albums were generally the worst in average, examining the best rated albums of all time in RYM reveals that there are many hip hop records amongst the chart [30]. The scope of this study limits the results to only be relevant to popular music, and they should not be generalized to represent the entire music landscape.

Another limitation was related to the analysis of genres. This is where we had to intentionally limit the scope of this study to only cover the topic from a very general standpoint. It would be interesting to cover each genre independently, and examine the birth and evolution of various subgenres. Such a study could also be accompanied by music information retrieval methods to identify the actual musical differences between the subgenres.

The results regarding the user ratings could also be distorted by survivorship bias, the idea that we only remember good things from the past, thus making older music seem better than it was. Our goal was to eliminate survivorship bias by covering all of the charting records at each time, but since we couldn't find the rating data for all of the albums, it could mean that only the best albums from the past were represented in our data. More recent albums also have a higher number of reviews on average, thus representing a larger amount of users. The limitations regarding data sourcing could be improved by the introduction of an official API, which would make the process more robust.

6 Conclusion

Our goal for this thesis was to gain understanding of the general dynamics of the popular music landscape, how innovations in music distribution and production tools have changed these dynamics and how music enthusiasts view these changes have affected the quality of music. We conducted our analysis using a data set of historical Billboard 200 charts and additional album level metadata from Rate Your Music.

In this thesis, We discovered that around the change of the 21st century, the popularity of CDs brought various changes to the dynamics of the chart. The sales of CDs made affordable high quality recordings available to the average consumer, increasing music sales and revenue (Figures 5.1 & 5.2), while also the amount of unique records increased on the chart (Figure 4.1). The increase in album variety also led to a larger portion of albums and artists appearing on the Billboard 200 for a shorter time. For example over half of the unique albums during 2000-2023 appeared on the list for less than one month, compared to just around 11% during 1977-2000.

During the 2010s, online music streaming services became the most common way of consuming music [1], bringing more changes in the dynamic of the charts. We noticed that the amount of yearly unique album appearances on the chart went down, and previous hit records that first appeared on the Billboard 200 chart over ten years ago started appearing on the chart again. This is evidence of the fact that

streaming has brought a way of profiting off older records, which now represents a large portion of overall music revenue [44].

In this thesis, we examined the online user ratings over time and discovered that around the late 1990s, there is a significant decrease in the average ratings of the charting albums. Our results show that this happened around the same time as the amount of albums started growing, possibly indicating that the increased variety brought by CD sales could have led into a decrease in album quality. The ratings are also not evenly distributed across various genres, with users generally being the most favorable towards rock albums, and most negative towards hip hop albums. We also visualized the popularity of each genre over time, most notably finding that since the late 2010s, hip hop has become the most popular genre in popular music. Based on our data, hip hop is also the first genre to pass both rock and pop in popularity.

References

- [1] IFPI, *Music consumer insight report*, 2018.
- [2] L. Colón, *Spotify*, Encyclopedia Britannica, Apr. 2024. [Online]. Available: <https://www.britannica.com/topic/Spotify> (visited on 05/02/2024).
- [3] A. Lerch, “The relation between music technology and music industry”, *Springer Handbook of Systematic Musicology*, pp. 899–909, 2018.
- [4] D. Arditi, “Digital subscriptions: The unending consumption of music in the digital era”, *Popular Music and Society*, vol. 41, no. 3, pp. 302–318, 2018.
- [5] H. Datta, G. Knox, and B. J. Bronnenberg, “Changing their tune: How consumers’ adoption of online streaming affects music consumption and discovery”, *Marketing Science*, vol. 37, no. 1, pp. 5–21, 2018.
- [6] P. Sherburne. “25 microgenres that (briefly) defined the last 25 years”. (2021), [Online]. Available: <https://pitchfork.com/features/lists-and-guides/microgenres-25th-anniversary/> (visited on 04/21/2024).
- [7] J. Fenney. “Is old music better?” (2015), [Online]. Available: <https://belwoodmusic.com/2015/06/21/is-old-music-better/> (visited on 05/02/2024).
- [8] A. Cross. “Must read: Why old music is better than today’s new music”. (2022), [Online]. Available: https://www.ajournalofmusicalthings.com/must-read-why-old-music-is-better-than-todays-new-music/?utm_content=cmp-true (visited on 05/02/2024).

-
- [9] S. Miller. “Are older music technologies better at capturing ‘liveness’ in recordings?” (2020), [Online]. Available: <https://www.thebritishacademy.ac.uk/blog/summer-showcase-2020-are-older-music-technologies-better-capturing-liveness-in-recordings/> (visited on 05/02/2024).
- [10] M. Katz, *Capturing sound: How technology has changed music*. Univ of California Press, 2010.
- [11] C. Gartenberg, “40 years ago, the sony walkman changed how we listen to music”, *The Verge*, 2019. [Online]. Available: <https://www.theverge.com/circuitbreaker/2019/7/1/20677636/sony-walkman-anniversary-tps-12-cassette-music-player-portable-mp3-evolution> (visited on 05/08/2024).
- [12] C. Shannon, “Communication in the presence of noise”, *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, Jan. 1949. DOI: 10.1109/jrproc.1949.232969. [Online]. Available: <https://doi.org/10.1109/jrproc.1949.232969>.
- [13] M. J. Benner and J. Waldfogel, “The song remains the same? technological change and positioning in the recorded music industry”, *Strategy Science*, vol. 1, no. 3, pp. 129–147, 2016. DOI: 10.1287/stsc.2016.0012.
- [14] R. Prey, “Nothing personal: Algorithmic individuation on music streaming platforms”, *Media, Culture & Society*, vol. 40, no. 7, pp. 1086–1100, 2018.
- [15] F. S. Barrett, K. J. Grimm, R. W. Robins, T. Wildschut, C. Sedikides, and P. Janata, “Music-evoked nostalgia: Affect, memory, and personality.”, *Emotion*, vol. 10, no. 3, p. 390, 2010.
- [16] M. D. Schulkind, L. K. Hennis, and D. C. Rubin, “Music, emotion, and autobiographical memory: They’re playing your song”, *Memory & Cognition*, vol. 27, pp. 948–955, 1999.

- [17] S. Laato and S. Rauti, “Comparing social media reactions to early game songs and final boss music in single player adventure games”, English, in *Proceedings of the 55th Annual Hawaii International Conference on System Sciences*, T. Bui, Ed., ser. Proceedings of the Annual Hawaii International Conference on System Sciences, jufoid=56894; Hawaii International Conference on System Sciences ; Conference date: 03-01-2022 Through 07-01-2022, Apr. 2022, pp. 3123–3132. DOI: 10.24251/HICSS.2022.384.
- [18] C. Davies, B. Page, C. Driesener, Z. Anesbury, S. Yang, and J. Bruwer, “The power of nostalgia: Age and preference for popular music”, *Marketing Letters*, vol. 33, no. 4, pp. 681–692, 2022.
- [19] M. B. Holbrook and R. M. Schindler, “Commentary on “is there a peak in popular music preference at a certain song-specific age? a replication of holbrook & schindler’s 1989 study””, *Musicae Scientiae*, vol. 17, no. 3, pp. 305–308, 2013.
- [20] C. Andrews, K. Gardiner, T. K. Jain, Y. Olomi, and A. C. North, “Culture, personal values, personality, uses of music, and musical taste.”, *Psychology of Aesthetics, Creativity, and the Arts*, vol. 16, no. 3, p. 468, 2022.
- [21] A. M. Petitbon and D. B. Hitchcock, “What kind of music do you like? a statistical analysis of music genre popularity over time.”, *Journal of Data Science*, vol. 20, no. 2, 2022.
- [22] J. Beckwith, “The evolution of music genre popularity”, Sep. 7, 2016. [Online]. Available: <https://thedataface.com/2016/09/culture/genre-lifecycles>.
- [23] M. Mauch, R. M. MacCallum, M. Levy, and A. M. Leroi, “The evolution of popular music: Usa 1960–2010”, *Royal Society open science*, vol. 2, no. 5, p. 150081, 2015.

- [24] F. Pachet, D. Cazaly, *et al.*, “A taxonomy of musical genres.”, in *RIAO*, Cite-seer, vol. 2, 2000, pp. 1238–1245.
- [25] J.-J. Aucouturier and F. Pachet, “Representing musical genre: A state of the art”, *Journal of new music research*, vol. 32, no. 1, pp. 83–93, 2003.
- [26] G. Tzanetakis and P. Cook, “Musical genre classification of audio signals”, *IEEE Transactions on speech and audio processing*, vol. 10, no. 5, pp. 293–302, 2002.
- [27] L. Wyse, “Audio spectrogram representations for processing with convolutional neural networks”, *arXiv preprint arXiv:1706.09559*, 2017.
- [28] H. Bahuleyan, “Music genre classification using machine learning techniques”, *arXiv preprint arXiv:1804.01149*, 2018.
- [29] S. Lippens, J.-P. Martens, and T. De Mulder, “A comparison of human and automatic musical genre classification”, in *2004 IEEE international conference on acoustics, speech, and signal processing*, IEEE, vol. 4, 2004, pp. iv–iv.
- [30] *Rate your music*, <https://rateyourmusic.com>. (visited on 12/12/2023).
- [31] *Album of the year*, <https://albumoftheyear.org>. (visited on 12/12/2023).
- [32] *Allmusic*, <https://allmusic.com>. (visited on 12/12/2023).
- [33] G. Trust. “How the charts have evolved across billboard’s history: The hot 100, billboard 200 & more”. (2019), [Online]. Available: <https://web.archive.org/web/20211123021748/https://www.billboard.com/pro/billboard-chart-history-evolution-milestones/> (visited on 11/23/2021).
- [34] Billboard. “Billboard hot 100 to include digital streams”. (Jul. 31, 2007), [Online]. Available: <https://web.archive.org/web/20130614022316/http://www.billboard.com/articles/news/1050326/billboard-hot-100-to-include-digital-streams> (visited on 01/29/2024).

- [35] “Billboard 200 makeover: Album chart to incorporate streams track sales”. (Nov. 19, 2014), [Online]. Available: <https://web.archive.org/web/20141122190912/http://www.billboard.com/articles/columns/chart-beat/6320099/billboard-200-makeover-streams-digital-tracks> (visited on 01/29/2024).
- [36] International Federation of the Phonographic Industry (IFPI). “Ifpi global music report: Global recorded music revenues grew 10.2% in 2023”. (2023), [Online]. Available: <https://www.ifpi.org/ifpi-global-music-report-global-recorded-music-revenues-grew-10-2-in-2023/> (visited on 04/21/2024).
- [37] M. Zipkin. “Best albums from the last decade, according to critics”. (2020), [Online]. Available: <https://stacker.com/music/best-albums-last-decade-according-critics> (visited on 04/21/2024).
- [38] *Similarweb*. [Online]. Available: <https://www.similarweb.com/website/rateyourmusic.com/#geography> (visited on 03/11/2024).
- [39] *Billboard-charts*. [Online]. Available: <https://github.com/guoguo12/billboard-charts>.
- [40] *Rymscraper*. [Online]. Available: <https://github.com/dbeley/rymscraper>.
- [41] V. Krotov and L. Silva, “Legality and ethics of web scraping”, 2018.
- [42] Flaky, *Rateyourmusic genre hierarchy for musicbee*, <https://gitlab.com/Flaky/mb-rym-hierarchy>.
- [43] *U.s. music revenue database*, <https://www.riaa.com/u-s-sales-database/>. (visited on 02/29/2024).
- [44] T. Gioia, “Is old music killing new music”, *The Atlantic*, vol. 23, 2022. [Online]. Available: <https://web.archive.org/web/20220207122546/https://www.>

theatlantic.com/ideas/archive/2022/01/old-music-killing-new-music/621339/.

- [45] E. A. Strobl and C. Tucker, “The dynamics of chart success in the uk pre-recorded popular music industry”, *Journal of Cultural Economics*, vol. 24, no. 2, pp. 113–134, 2000.
- [46] B. Duinker, “Song form and the mainstreaming of hip-hop music”, 2020.
- [47] K. Sanneh, “The rap against rockism”, *The New York Times*, vol. 31, pp. 351–4, 2004.

Appendix A

Listing 1 Algorithm used in this study to compare artist and album names in the two different data sources and check for possible abbreviations

```
def is_abbrev(abbrev, text):
    abbrev=abbrev.lower()
    text=text.lower()
    words=text.split()
    if not abbrev:
        return True
    if abbrev and not text:
        return False
    if abbrev[0]!=text[0]:
        return False
    else:
        return (is_abbrev(abbrev[1:], '_' .join(words[1:])) or
                any(is_abbrev(abbrev[1:], text[i+1:])
                    for i in range(len(words[0]))))
```

Artist	Mean peak album position	Album count
Drake	1.714286	14
Eminem	1.750000	12
Taylor Swift	6.263158	19
Madonna	12.884615	26
Kenny Chesney	14.571429	21
JAY-Z	14.666667	18
Bruce Springsteen	23.294118	34
Barbra Streisand	23.920635	63
The Rolling Stones	26.379310	58
The Beatles	37.896552	58

Table A.1: Mean peak album positions for the artists with most top 1 albums

Artist	Title	first date	last date	Years
The Beatles	Revolver	1966-09-03	2023-01-07	56.4
The Rolling Stones	Beggars Banquet	1968-12-14	2023-05-06	54.4
The Beatles	Abbey Road	1969-10-18	2023-07-01	53.7
The Beatles	Sgt. Pepper's Lonely Hearts Club Band	1967-06-24	2021-01-16	53.6
The Who	The Who Sell Out	1968-01-06	2021-05-08	53.4
Crosby, Stills, Nash & Young	Deja Vu	1970-04-04	2023-02-04	52.9
The Who	Who's Next	1971-08-14	2023-09-30	52.2
The Beatles	Let It Be (Soundtrack)	1970-05-30	2022-02-26	51.8
The Rolling Stones	Hot Rocks 1964-1971	1972-01-08	2023-09-23	51.7
The Jimi Hendrix Experience	Smash Hits	1969-08-02	2020-11-28	51.4

Table A.2: Albums that have appeared on the chart for the longest after the first appearance

Title	Year	Weeks on list
1	2000	540
Abbey Road	1969	490
Sgt. Pepper's Lonely Hearts Club Band	1967	233
The Beatles [White Album]	1968	215
The Beatles 1967-1970	1973	188
Revolver	1966	94
Magical Mystery Tour (Soundtrack)	1967	93
Let It Be (Soundtrack)	1970	79
Meet The Beatles!	1964	74
Beatles '65	1965	71
Rubber Soul	1965	70
A Hard Day's Night (Soundtrack)	1964	56
The Beatles' Second Album	1964	55
Help! (Soundtrack)	1965	46
Beatles VI	1965	41
Anthology 2	1996	37
Yesterday And Today	1966	32
Anthology 1	1995	29
Anthology 3	1996	16

Table A.3: Albums by The Beatles that have achieved the top 1 spot on Billboard 200

Title	Year	Weeks	Peak position
Moana	2016	352	2
The Sound Of Music	1965	238	1
The Greatest Showman	2017	191	1
Frozen	2013	179	1
Beaches	1989	176	2
The Big Chill	1983	161	17
Doctor Zhivago	1966	157	1
Guardians Of The Galaxy: Awesome Mix Vol. 1	2014	147	1
Saturday Night Fever	1976	137	1
Pitch Perfect	2012	126	3

Table A.4: Top soundtracks by weeks on chart

Title	Year	Weeks	Peak position
ESPN Presents: Jock Jams Volume 1	1995	105	30
Jesus Christ Superstar	1970	102	1
Dreamville & J. Cole: Revenge Of The Dreamers III	2019	98	1
WoW Worship: Today's 30 Most Powerful Worship Songs	1999	89	70
Songs 4 Worship: Shout To The Lord	2001	80	51
Jock Jams Vol. 2	1996	63	10
Super Hits	1967	60	12
Monsters Of Rock	1998	58	112
Free To Be...You And Me	1973	58	68
Pure Disco	1996	57	83

Table A.5: Top compilations by weeks on chart

Artist	Title	Weeks
Original London Cast Recording	The Phantom Of The Opera: Highlights	331
Enigma	MCMXC A.D.	282
Metallica	Metallica	281
Original London Cast Recording	The Phantom Of The Opera	255
Nirvana	Nevermind	252
Pearl Jam	Ten	250
Van Morrison	The Best Of Van Morrison	242
Enya	Shepherd Moons	238
Queen	Greatest Hits	233
Garth Brooks	Garth Brooks	224

Table A.6: Top 10 albums with most chart weeks in 1977-2000

Artist	Title	Weeks
Eminem	Curtain Call: The Hits	662
Creedence Clearwater Revival	Chronicle: The 20 Greatest Hits	658
Bruno Mars	Doo-Wops & Hooligans	655
Guns N' Roses	Greatest Hits	654
Adele	21	590
Kendrick Lamar	good kid, m.A.A.d city	581
Drake	Take Care	563
The Beatles	1	539
Lana Del Rey	Born To Die	515
Drake	Nothing Was The Same	493

Table A.7: Top 10 albums with most chart weeks since 2000

Artist	weeks
Alabama	708
Metallica	645
Madonna	644
George Strait	620
Billy Joel	596
Van Halen	596
Elton John	592
Kenny G	570
Rod Stewart	568
Aerosmith	558

Table A.8: Artists with the most chartlisted weeks in 1977-2000

Artist	weeks
Eminem	1045
Taylor Swift	863
Tim McGraw	862
Rihanna	788
Kidz Bop Kids	781
Kenny Chesney	770
George Strait	750
Nickelback	739
Jason Aldean	736
Drake	728

Table A.9: Artist with the most chartlisted weeks since 2000