What Do We Listen To Today?
An Analysis of Popularity in Music Streaming

Yonghyun Kang
Purdue University
kang612@purdue.edu

Kevin Huang
Purdue University
huan2151@purdue.edu

Frederick Lu
Purdue University
lu1137@purdue.edu

***Abstract*—The transition of music consumption from offline to online realm, facilitated by technological advancements, has fundamentally transformed the music distribution landscape. In the streaming era, platforms that utilize data-driven techniques, such as tracking user behavior and analyzing audio data, have profoundly impacted on music consumption trends. As a result, technologies like recommender systems integrated into music streaming platforms contribute to bringing large, diverse audiences together, shaping a unified standard of popular music. This report presents features and patterns commonly observed among trending music based on such newfound standards. Our work particularly focuses on quantitatively examining listening trends and musical characteristics of popular music by employing exploratory data analysis on datasets founded upon the library from Spotify, spanning from 2010 to 2023. By analyzing these trends, we aim to highlight how data-driven methodologies are influencing the crowds' music preferences and consumption dynamics in the digital age.**

1. Introduction

As a result of technological advancement, modern web-based platforms such as social networking services have significantly transformed the way we listen to music in recent years. In 2024, music streaming services stake 84% of the total United States's pre-recorded music industry revenue, demonstrating their predominant position in the music industry [1]. Not only vast volumes of music libraries but personalized recommendation systems and statistical reports they provide (e.g., Spotify's Year Wrapped) actively engage in the social aspect of consumers' listening behaviors. Spotify, one of the largest streaming platforms in the global market, shaped the consumers' music discovery process and gained majority by leveraging technology such as "advanced search engines" and "algorithms" on their music recommendation systems [2].

Today, streaming platforms and social network services (e.g., YouTube and TikTok) are continuously increasing their influences over daily lives, specifically bringing social and interactive aspects [3], and music played on those platforms as itself or as background music for short-form videos become able to be popularized among niche groups or more larger heterogeneous groups through the sharing process so-called "viral". The transformation of the music industry as data-driven businesses, turning music listening into measurable behavior, contributes to constantly evolving recommendation systems and reshaping user preferences as well [4].

In this study, we investigate features observed in music listening trends in the digital era. Particularly, we quantify and examine past Spotify streaming records in order to reveal features of popular music that affected by data-driven recommendation systems. By employing exploratory data analysis, we extend our work to future predictions or explanations including visualization process.

2. Approach and Methodology

This study aims to explore different aspects of information related with popular music among the crowds and to identify their repeated or common characteristics. In this work, our approach involves employing multiple quantitative analysis techniques, emphasizing both descriptive and inferential statistical methods to uncover features or correlations. Mainly, this work is driven by following objectives:

a) To understand the motion of popularity, i.e., to track the transition of popular songs' characteristics over time.
b) To investigate the correlation between musical characteristics and popularity, particularly we focus on analyzing audio features commonly associated with top-charting songs.
c) To model the predictive relationship between musical characteristics and streams.

For the quantitative data analysis, we choose Spotify, particularly, as our primary source of datasets for two major reasons. First, we value the demographic broadness in order to minimize population-specific listening biases reflected on top charts. As of the third quarter of 2023, Spotify holds a leading position in the global music streaming market with a 31.7% share of subscribers [5]. Therefore, we can infer that such large portion may possible comprises large, scatter demographics in different range of biological and cultural differences. Second, due to the limited time and computational power allowed to this study, we decide to focus on platforms who are open to providing streaming data and each track's meta information, including musical characteristics, as well. Spotify is providing pre-evaluated audio features in their API, such as acousticness and danceability, and this is significantly essential for this work to be done by reducing computation time spending on analyzing each music's audio characteristics. Examples of audio features provided by Spotify is described in Table 1.

As our study constrained in time, we utilize music datasets from Kaggle instead of collecting data through Spotify's public API. Consequently, each selected dataset provide only partial information for each song relevant to this study. Hence, it is crucial to preprocess multiple datasets into a reliable and comprehensive format. The numbers of records of each dataset we used in this study are vary from 165K to 4.6K. Thus, in order to filling missing values occurred by merging multiple datasets into the master-data table, we impute the missing values with equivalent columns' mean values in some data wrangling process. Additionally, we focus on stream data represent global trending and specific countries, Canada, France, Germany, Japan, South Korea, United Kingdom, and United States. By omitting other countries, we are able to process large portion of data faster and facilitate exploring noteworthy information easier. The entire workflow, including code, analysis process, and visualized charts, is pushed on our public GitHub repository[1].

2.1. Datasets

In this work, we utilize two distinct groups of public datasets, each serving for a specific purpose and complementing the other in providing a comprehensive analysis. The first group consists of datasets that capture the statistical trends of music over a certain period of time (from 2010 to 2023), such as number of streams or popularity of each track calculated by Spotify. These datasets provide insights into platform-specific trends, seasonal variations, and user preferences, helping us understand how external factors influence music popularity. The second group includes datasets containing audio features mentioned above, including acousticness, danceability, duration, energy, instrumentalness, liveness, loudness, speechiness, tempo, and valence. Description of each is as follows [7]:

- **acousticness**: "A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic."
- **danceability**: "Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable."
- **duration**: "The duration of the track in milliseconds."
- **energy**: "Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy."
- **instrumentalness**: "Predicts whether a track contains no vocals. 'Ooh' and 'aah' sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly 'vocal'. The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0."
- **liveness**: "Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live."
- **loudness**: "The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typically range between -60 and 0 db."
- **speechiness**: "Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks."
- **tempo**: "The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration."
- **valence**: "A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry)."

By analyzing these two groups of datasets together, we can capture a diverse aspects of user behavior, from platform-specific or seasonal trends to the inherent properties of the music itself. This approach allows us to comprehend listening tendencies across different time thoroughly.

2.1.1. List of datasets used in this study.

- **Spotify Charts**: The Spotify charts dataset has more than 160K records from 2016 to 2021. The data in this set, including top 200 music of multiple

1. github.com/cs176-2024-fa-kda/notebook

TABLE 1. SPOTIFY TRACKS DATASET [6]

track_id	track_name	popularity	danceability	energy	loudness	speechiness	···
5SuOikwiRyPMVoIQDJUgSV	Comedy	73	0.676	0.461	-6.746	0.143	···
4qPNDBW1i3p13qLCt0Ki3A	Ghost - Acoustic	55	0.42	0.166	-17.235	0.0763	···
1iJBSr7s7jYXzM8EGcbK5b	To Begin Again	57	0.438	0.359	-9.734	0.0557	···
6lfxq3CG4xtTiEg7opyCyx	Can't Help Falling In Love	71	0.266	0.0596	-18.515	0.0363	···
5vjLSffimiIP26QG5WcN2K	Hold On	82	0.618	0.443	-9.681	0.0526	···
		⋮					

country, help us to track regional trends specifically on Spotify [8].

- **Top Streamed Spotify Songs by year 2010 – 2023**: The top streamed Spotify songs by year 2010 – 2023 dataset also provides long period of historical streaming data of Spotify from 2010 to 2023 [9].
- **Spotify Tracks Dataset**: The Spotify tracks dataset contains multiple meta information for individual music available on Spotify, which can be utilized to relate each song's musical characteristics [6].
- **Most Streamed Spotify Songs 2024**: The most streamed Spotify songs 2024 dataset presents the most recent statistical numbers of each music in multiple platforms, such as YouTube and TikTok. By aggregating with other datasets, we can extend our understanding into different platforms with recent data points and ensure the reliability of our findings by validating them [10].

3. Limitations

This study has several limitations. First, due to the heavy reliance on top trending tracks, a contrast between trending and non-trending tracks cannot be rendered in our interpretations. This is a significant limitation, as it may lead a partial biased finding as a general trend. Hence, we suggest that future research incorporates a more random sample that includes both trending and non-trending tracks in order to provide more comprehensive understanding in music listening trends across broad demographics. Additionally, since this work does not covers other platforms' streaming data but only Spotify, it is difficult to extend our findings into the entire populace. We believe that other platforms, specifically social networking services, such as Instagram and TikTok, may play significant role in creating trending and influencing form of standard in popular music. Additional study should examine more data across multiple demographics in different platforms.

Second, this work does not account for individual listening behaviors or trends commonly observed among specific demographic subgroups (e.g., age or gender), but rather considers crowd as a whole. As a result, our findings may overlook important differences in musical preferences and trends across various groups. For example, younger listeners may gravitate toward genres that have higher danceability, while older listeners may prefer less louder songs. Future research should delve into these subgroup dynamics in more detail, examining how demographic factors influence listening trends.

Third, since our work is focused on quantitative measurement and analysis, the influence of cultural, linguistic, or societal background in music selection (i.e., a requiem for Christians), which we believe they contributes deeply, is also not reflected in the final interpretation. Moreover, even our work is based on a platform that is considered as having broad spectrum of demographics and *unbiased*, it does not represents the entire crowd listening music in digital forms, and thus, our conclusion cannot be generalized into the behavior of total consumers in music streaming market.

4. Analysis and Results

4.1. Streaming trends over time

4.1.1. Impact of the pandemic in a global music consumption and emerging trends. With the growth of the music industry that propelled by advanced streaming platforms, such as Apple Music, Spotify, and YouTube Music, the number of total music streaming of Spotify users in each year gradually become increasing for the past decade [11]. The Figure 1 reveals a notable surge in the number of streams in 2020, followed by a steep decline in 2021, even though the total number of tracks played remained relatively stable increase during this period. This pattern suggests that an external global-scale event, such as the COVID-19, may have played a significant role in this fluctuation. Specifically, we hypothesize that the number of streams could be closely related with the quarantine policies implemented worldwide in 2020 in response to the pandemic.

Yeung (2020) supports this notion, suggesting that a positive relationship between lockdowns and increased music consumption during the same period. The research also highlights a rise in *nostalgic* music in 2020, attributing this shift to the emotional impact of the lockdown and the role of music in coping with isolation [12]. After that, as the pandemic crisis stabilizes and lockdowns are eased, we infer that consumer behavior may show a correlation with the general shift back to offline consumption patterns [13].

4.1.2. Evolution of genre preferences and trends over time. The evolution of genre preferences over time can provide insights into shifting cultural dynamics and changing consumer behaviors in music consumption. Figure 2 shows the total number of streams for tracks from each root-genre,

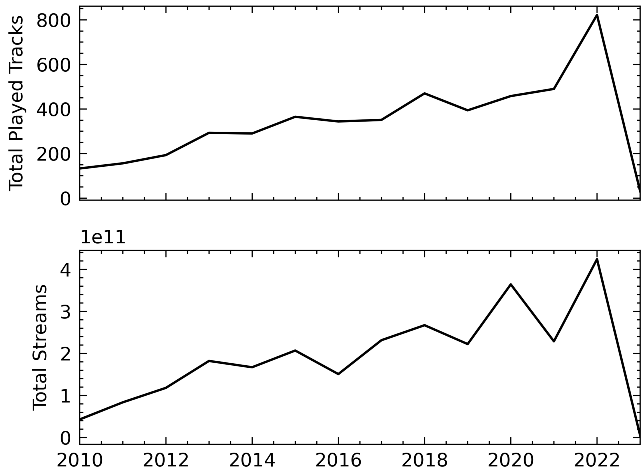

Figure 1. Number of the most popular songs' total global streaming over years from 2010 to 2023. The upper figure displays the total number of unique tracks ranked as the most popular each year from 2010 to 2023. The lower figure illustrates the total number of streams for these tracks over the same period.

accumulated in their year of release between 2010 and 2023. This indicates consistent trends, with Pop and Hip-Hop/Rap being the most dominant genres on Spotify for a long period of time.

This consistency is likely driven by the broad appeal and adaptability of these genres across diverse demographics. The familiarity that listeners have with these genres, combined with their pervasive presence in mainstream media, particularly social networking services, has enabled Pop and Hip-Hop/Rap to reinforce their dominance.

This dominance also suggests that these genres as the most popular lays the groundwork for further research into the specific characteristics that make songs within these categories resonate with audiences. For instance, analyzing audio features such as danceability, energy, tempo, and lyrical themes in Pop and Hip-Hop/Rap could provide insights into the elements that drive their success.

4.2. Musical characteristics of popular tracks

4.2.1. Relationship between genres and popularities of music. As music evolve rapidly, they constantly evolve and become complex by blending with multiple genres, resulting in unique and diverse sounds. According to Silver (2016), music that incorporate "complexity" and multiple genres tend to be more popular rather than monolithic, single-genre tracks [14]. Hence, we explore patterns of each genre's occurrences in popular tracks. At first, we delve into co-occurrences of each genre by performing PCA based on a p-by-p matrix, where p is the number of unique genres in our datasets and M_{ij} represents the frequency of genre i with genre j.

Figure 3 shows the co-occurrence relationships among 1,128 unique genres analyzed in our datasets. Notably, two genres 'pop' and 'rap', exhibit distinct isolation from other data points. This observation suggests that, unlike most genres, 'pop' and 'rap' co-occur with a broader range of other genres, indicating lower cohesion within specific genre clusters. Consequently, these genres demonstrate a more generalized and versatile nature compared to others, which tend to exhibit tighter clustering due to their niche or specialized appeal. Thus, we can infer that such genres are more prevalent and ubiquitous in the music industry, reflecting their capacity to be blended with a wider spectrum of musical categories, reinforcing their dominance in mainstream culture.

Accordingly, Figure 4 shows that 'pop' has the most subcategories, followed by 'rap', suggesting that these genres are the most popular and widely intersected among the others.

Likewise, Figure 5 illustrates the distribution of preferences across different pop sub-genres. In the graph, dance pop stands out as the dominant sub-genre, capturing the largest share with 17.6%. This highlights that even at the sub-genre level, a dominant taste prevails, suggesting that the musical characteristics of a particular sub-genre are likely to influence or be associated with broader trends in overall music popularity.

When we extend our analysis beyond the 'pop', we observe the popularity of other genres based on their total streams. Figure 6 highlights other root genres, such as 'rock', 'rap', and 'edm', dominating the top rankings. Interestingly, sub-genres like 'modern rock' and 'urbano latino' rank higher than these root genres. This suggests a global shift in listener preferences toward sub-genres that offer complex, innovative sounds, demonstrating their growing influence in shaping modern music trends and audience tastes.

4.2.2. Audio features among popular music. To understand the features of popular tracks' audio data, we analyzes pre-evaluated values from Spotify by statistical manner. Figure 7 shows how normalized values (from 0 to 1) of each category of audio features are generally clustered across tracks. This provides a visual representation of common trends of popular songs' audio feature values. The thickness in each column indicates how many tracks have value in that feature value. The noticeable thickness in a column provides a visual representation of the value's prevalence among the group, i.e., as it becomes thicker it appears in the data points of the group of set more frequently. At a glance, assuming that the value of each category is evenly distributed, we can infer that most popular songs has high danceability while its length is relative shorter than the entire scale. Additionally, the duration of these tracks is relatively shorter compared to the full scale of possible values, suggesting that shorter tracks are more common in the dataset. This analysis highlights the general tendencies in popular music, though there are still variations across different tracks.

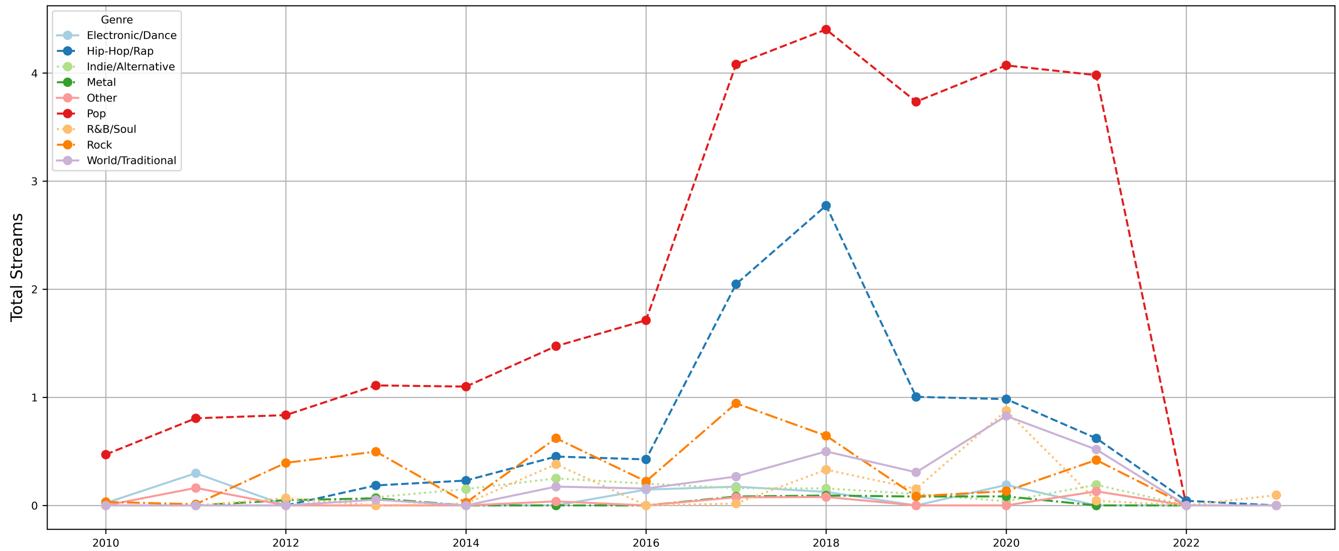

Figure 2. Total annual number of streams across different genres

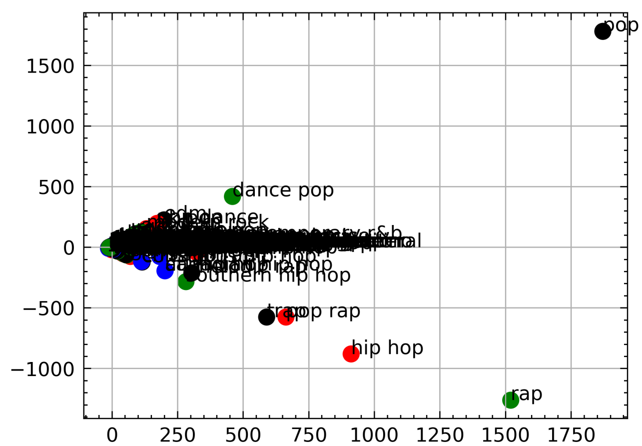

Figure 3. Each genre's co-occurrences with other genres among popular music in Spotify from 2010 to 2024.

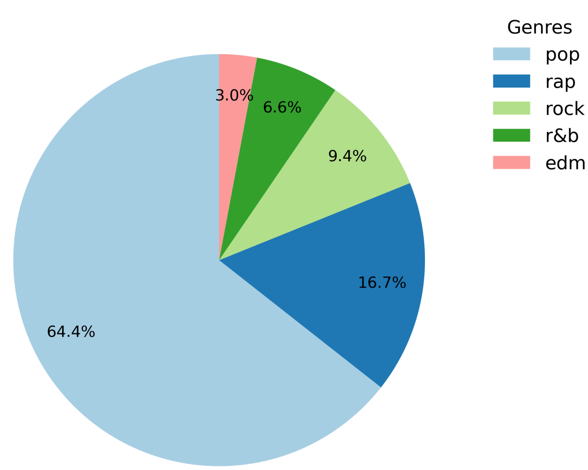

Figure 4. Proportion of total subcategories of each *global* genre

5. Conclusion

In this study, we have analyzed various aspects of features observed among popular songs on Spotify from 2010 to 2023, focusing on how data-driven methodologies influence music consumption and preference trends in the digital age. Our exploration of music metadata and listening patterns has revealed notable trends in popular music, such as recurring audio characteristics that resonate with a broad audience, including tempo, danceability, and energy levels. These findings underline the significant role of recommender systems in shaping listening behavior by promoting songs that align with the identified preferences of users.

Since our study is confined to the limited records of streaming data and music metadata from a specific platform and over the short term, we suggest that further extensive analysis based on a larger sample size is necessary to conclude the comprehensive understanding of listening tendency in more multi-layered groups. Specifically, future studies should account for other variables, such as biological (e.g., affection from certain types of hormones), societal (e.g., generation-wide tendency), cultural (e.g., conformity effect among youths or people from collectivistic culture) or religious (e.g., prohibition of music containing explicit language in certain religions) background's influence on listening behavior, to elucidate the underlying correlation of such factors and listening patterns in a particular group. Accordingly, we encourage further investigation into how dynamic audio features, such as hooks or patterns within a song's progression, contribute to listener engagement and song popularity. Such analyses would provide deeper insights into the intricate relationship between music structure

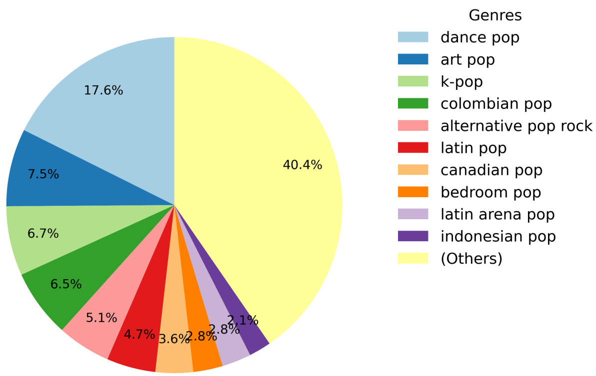

Figure 5. Share of top 10 sub-*pop* genres based on the number of streams of each genre from 2010 to 2023. Note: The figure is generated from the total 126 unique sub-pop genres.

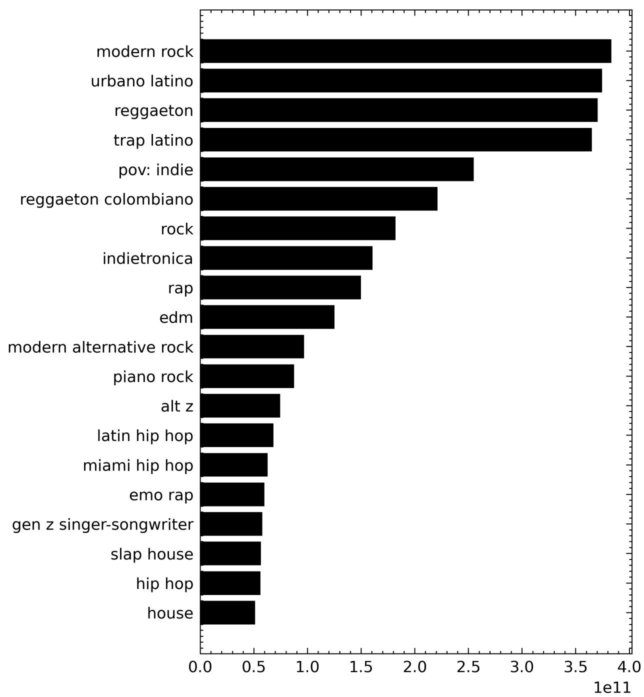

Figure 6. Most popular genres except pop through 2010 to 2023. The upper figure shows most popular ten genres except pop based on the sum of individual tracks' popularity value. The lower figure shows most streamed ten genres except pop based on the sum of individual tracks' number of stream.

and its reception, ultimately enhancing our understanding of the data-driven forces shaping modern music trends.

References

[1] M. Bass, "2024 mid-year music industry revenue report," The Recording Industry Association of America, Tech. Rep., 2024. [Online]. Available: https://www.riaa.com/wp-content/uploads/2024/08/RIAA-Mid-Year-2024-Revenue-Report.pdf

[2] Y. Kjus, "Musical exploration via streaming services: The norwegian experience," *Popular Communication*, vol. 14, no. 3, pp. 127–136, 2016. [Online]. Available: https://doi.org/10.1080/15405702.2016.1193183

[3] D. Lee, J. Yejean Park, J. Kim, J. Kim, and J. Moon, "Understanding music sharing behaviour on social network services," *Online Information Review*, vol. 35, no. 5, pp. 716–733, jan 2011. [Online]. Available: https://doi.org/10.1108/14684521111176462

[4] E. DROTT, "Why the next song matters: Streaming, recommendation, scarcity," *Twentieth-Century Music*, vol. 15, no. 3, pp. 325–357, 2018.

[5] "Music streaming services subscribers market shares 2023 — Statista — statista.com," https://www.statista.com/statistics/653926/music-streaming-service-subscriber-share/?utm_source=chatgpt.com, [Accessed 08-12-2024].

[6] M. Pandya, "Spotify tracks dataset," 2022. [Online]. Available: https://www.kaggle.com/dsv/4372070

[7] "Web API Reference — Spotify for Developers — developer.spotify.com," https://developer.spotify.com/documentation/web-api/reference/get-audio-features, [Accessed 04-12-2024].

[8] D. Dave, "Spotify charts," 2021. [Online]. Available: https://www.kaggle.com/ds/1265407

[9] I. Tokarchuk, "Top streamed spotify songs by year 2010 – 2023," 2024. [Online]. Available: https://www.kaggle.com/datasets/irynatokarchuk/top-streamed-spotify-songs-by-year-2010-2023

[10] N. Elgiriyewithana, "Most streamed spotify songs 2024," 2024. [Online]. Available: https://www.kaggle.com/dsv/8700156

[11] R. A. Rahimi and K.-H. Park, "A comparative study of internet architecture and applications of online music streaming services: The impact on the global music industry growth," in *2020 8th International Conference on Information and Communication Technology (ICoICT)*. IEEE, 2020, pp. 1–6.

[12] T. Y.-C. Yeung, "Did the covid-19 pandemic trigger nostalgia? evidence of music consumption on spotify," *Evidence of Music Consumption on Spotify (August 21, 2020)*, 2020.

[13] K. L. Jensen, J. Yenerall, X. Chen, and T. E. Yu, "Us consumers' online shopping behaviors and intentions during and after the covid-19 pandemic," *Journal of Agricultural and Applied Economics*, vol. 53, no. 3, pp. 416–434, 2021.

[14] D. Silver, M. Lee, and C. C. Childress, "Genre complexes in popular music," *PloS one*, vol. 11, no. 5, p. e0155471, 2016.

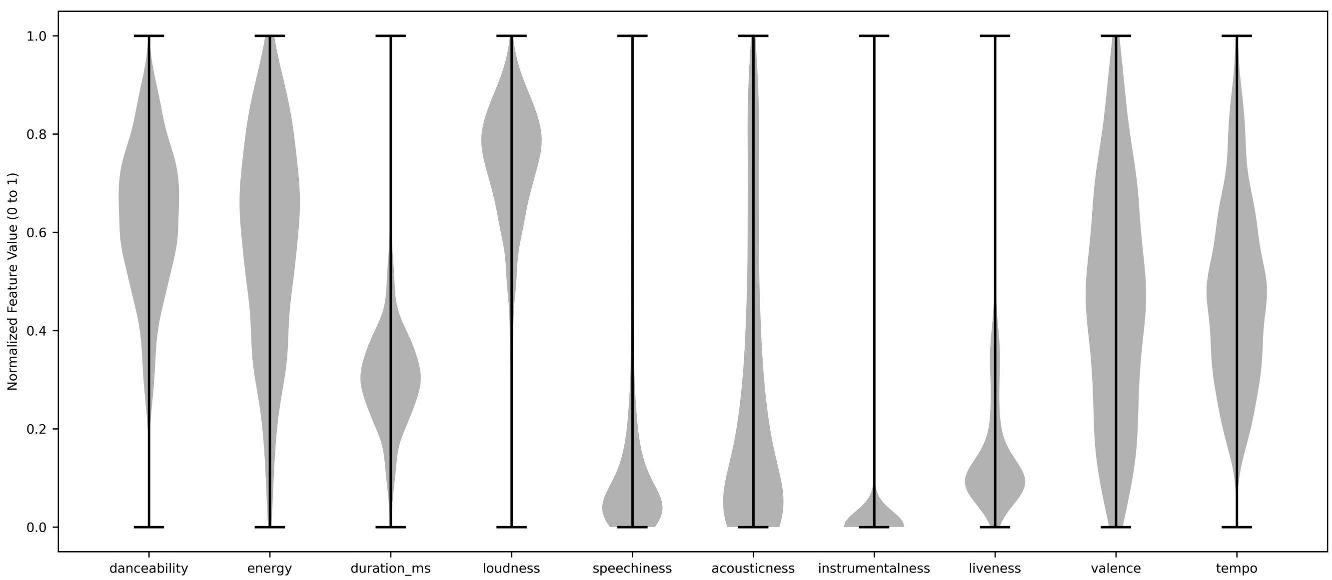

Figure 7. Audio attributes of the total tracks' values expressed as normalized between 0 to 1 regardless genres.