THE LISTENBRAINZ LISTENS DATASET

Kartik Ohri
MetaBrainz Foundation Inc.
lucifer@metabrainz.org

Robert Kaye
MetaBrainz Foundation Inc.
rob@metabrainz.org

ABSTRACT

The ListenBrainz listens dataset is a continually evolving repository of music listening history events submitted by all ListenBrainz users. Currently totalling over 800 million entries, each datum within the dataset encapsulates a timestamp, a pseudonymous user identifier, track metadata, and optionally MusicBrainz identifiers facilitating seamless linkage to external resources and datasets. This paper discusses the process of raw data acquisition, the subsequent steps of data synthesis and cleaning, the comprehensive contents of the refined dataset, and the diverse potential applications of this invaluable resource. Although not the largest dataset in terms of music listening events (yet), its distinctiveness lies in its perpetual evolution, with users contributing data daily. This paper underscores the significance of the ListenBrainz listens dataset as a significant asset for researchers and practitioners alike, offering insights into music consumption patterns, user preferences, and avenues for further exploration in the fields of music information retrieval and recommendation systems.

Keywords: novel datasets, digital archives, metadata, linked data

1. INTRODUCTION

The advent of digital music streaming has led to an explosion of data on user listening habits. As the most prevalent form of music consumption today, with streaming accounting for 84% of total U.S. music revenue in 2023 [1] , this data holds immense potential for understanding trends, developing recommendation systems, and personalizing the user experience. However, most of this data is locked within commercial platforms and inaccessible to researchers or the public [1]. This lack of transparency hinders open-source development and independent research efforts in the music information retrieval field. AI-driven music recommendation systems, personalized playlists, and even music generation algorithms rely heavily on vast datasets of user behavior to function effectively [2]. Open access to music consumption habits datasets is crucial in ensuring that these algorithms are developed and trained in a manner that is fair, transparent, unbiased, and representative of diverse musical tastes.

This paper introduces the ListenBrainz listens dataset, explores its contents, and potential applications. We will discuss the unique characteristics that distinguish it from other music datasets and highlight its significance as a valuable resource for researchers, practitioners, and music enthusiasts alike. Our goal is to provide the research community with a valuable resource for analyzing evolving music consumption patterns, exploring user preferences, and advancing open-source music information retrieval systems.

2. RELATED WORK

A few public music listening history datasets exist, most built upon data extracted from the social music platform Last.fm. These include the Last.fm Dataset-360K [5]; the Last.fm Dataset-1K [5], the LFM-1B dataset [6] and the LFM-2B dataset [7]. The LFM-1B dataset [2] and the LFM-2B dataset [3] are not available anymore due to licensing issues.

All of these datasets were superseded by the introduction of the Music Listening History Dataset (MLHD) in 2017. MLHD stands out as one of the largest and most comprehensive publicly available datasets of music listening histories even today. It contains over 27 billion timestamped listening events from 583,000 users, enriched with demographic information and MusicBrainz identifiers for linking with external resources [3]. MLHD has been extensively used in research on music recommendation, user behavior analysis, and temporal trends in music consumption. To our knowledge, no newer datasets of comparable size and scope surpassing MLHD have been released since, highlighting the continued relevance and value of this resource. However, it is no longer possible to update MLHD with new data from Last.fm as the API endpoints originally used to curate the dataset have now been taken down [8].

The Music Streaming Sessions Dataset (MSSD), unveiled by Spotify, takes a unique approach by centering on listening sessions rather than individual track plays. It encompasses 160 million sessions, each providing in-

[1] U.S. Recorded Music Revenues Data by Format taken from the RIAA U.S. Music Revenue Database https://www.riaa.com/u-s-sales-database/

[2] Hosting page for LFM-1B dataset http://www.cp.jku.at/datasets/LFM-1b/

[3] Hosting page for LFM-2B dataset http://www.cp.jku.at/datasets/LFM-2b/

Feature	MLHD [3]	MSSD [4]	ListenBrainz
Source	Last.fm Scrobbles	Spotify Streaming Logs	ListenBrainz User Submissions
Size	27 Billion Listening Events	160 Million Listening Sessions	800+ Million (and growing) Listening Events
Scope	Individual Track Plays	Listening Sessions (upto 20 tracks)	Individual Track Plays
Content	Timestamp, Basic Track Metadata, Limited MBIDs, User Demographics	Timestamp, User Actions, Track Metadata, Audio Features, Playlist Snapshots	Timestamp, Extended Track Metadata, Comprehensive MBIDs Links
Updates	Static (Last Updated 2017)	Static (Last Updated 2019)	Dynamic (Continuously Updated)
Strengths	Large size, Comprehensive user demographics, MBIDs for linking	Focus on listening sessions, Includes audio features, Counterfactual evaluation subset	Continuously updated, User-controlled data, Diverse data sources (streaming, local files), Extended Metadata Coverage, MBIDs for linking

Table 1: Comparison of the important music listening datasets

sights into user actions within the session, audio features of the tracks, and corresponding track metadata [4]. While MSSD offers valuable data for analyzing the dynamics of listening sessions, its scope is more confined compared to MLHD. MSSD encompasses a smaller user base and covers a shorter time frame. As of today, the MSSD dataset is not available for download publicly [4].

A common limitation shared by all the mentioned datasets, including both MLHD and MSSD, is their static nature. They represent snapshots of data frozen at a particular moment in time, lacking updates since their initial release. This inherent static nature raises concerns about their ability to accurately reflect contemporary music consumption patterns and trends. Furthermore, these datasets are missing data on music released after their creation, potentially restricting their usefulness for research inquiries focused on recent musical trends and user preferences. Additionally, the track metadata provided in MLHD and MSSD is limited to basic information such as artist, track, and album names. In contrast, ListenBrainz allows users to submit any additional metadata they deem relevant alongside their listening events, providing a richer and more comprehensive dataset for analysis.

Table 1 offers a concise overview of the key characteristics and differences between the Music Listening History Dataset, the Music Streaming Sessions Dataset, and the ListenBrainz Listens Dataset.

3. BACKGROUND

3.1 Music Listening History

Music listening histories serve as extensive timelines of an individual's music consumption, offering valuable insights into their preferences, habits, and evolving tastes. Aggregating these histories across different timeframes uncovers broader patterns and trends in listening behavior [9] [10]. The open availability of such data holds immense potential for advancing music information retrieval research, enhancing recommendation systems, and fostering a deeper understanding of the relationship between individuals and music.

3.2 Data Donation

Data donation is a method of data collection which typically involves users proactively sharing their digital trace data, often by requesting and exporting their data from online platforms, with researchers [11]. Data donations are commonly used in the field of communications, especial social media, research [12]. The usefulness of data donations in music research being increasingly recognized as exemplified by the Fair Muse project [13].

3.3 The ListenBrainz Project

The ListenBrainz listens dataset has been developed as a part of the broader open source project, ListenBrainz [5]. The project is maintained by the MetaBrainz Foundation, a non-profit organization dedicated to promoting open data initiatives in the music domain. The organization is renowned for its over two-decade-long stewardship of the comprehensive free and open source MusicBrainz database [6]. All ListenBrainz data is generously licensed under the CC0 license, granting unrestricted use and creating a collaborative environment for research and development.

4. THE LISTENBRAINZ LISTENS DATASET

4.1 Data Collection

The ListenBrainz dataset is entirely crowdsourced, with users actively contributing their listening histories. Lis-

[4] MSSD Dataset downlad page https://www.aicrowd.com/challenges/spotify-sequential-skip-prediction-challenge

[5] https://listenbrainz.org/

[6] History and deatils of the MetaBrainz Foundation Inc. and the MusicBrainz project can be found at https://metabrainz.org/about

	Traditional Data Donation	ListenBrainz
Data Acquisition	Users request data from platform and donate to researchers.	Data submitted directly to ListenBrainz (automatically or manually).
Temporality	One-time or infrequent bulk data donations.	Continuous, regular data contribution.
User Effort	Active user involvement required in export and donation	Minimizes user effort after initial setup.
Data Scope	Limited to a single platform or service.	Aggregates data from multiple sources.
User Control	Limited control post data donation.	Offers ongoing user control (editing, deletion, or contribution cessation).
Data Utilization	Often for specific research projects with limited broader application.	Continuously growing, multi-purpose dataset for diverse research and the music community.

Table 2: Data Collection: Traditional Data Donation vs. ListenBrainz Approach

tenBrainz's data collection approach shares its ethos with traditional data donation approaches. Both involve voluntary participation and aim to provide transparency regarding data usage.

However, the traditional data donation approach has some limitations. The donated data is retrospective and represents a one-time export. Repeated donations require users to navigate potentially complex processes which discourage participation [12]. To overcome these limitations, ListenBrainz provides multiple ways for users to setup automatic submission of listening events from their music streaming platforms and local music players on a continuous basis. Table 2 sums up the differences between the traditional and ListenBrainz approach.

Users can submit their data through various methods.

1. APIs and local media players: ListenBrainz provides a free and open API [7] allowing manual submission of listening histories and facilitates the development of plugins for music players, automating the process for seamless and reliable data collection [8] . There is a Last.fm compatible API available as well which allows existing Last.fm clients to readily integrate with ListenBrainz [9] .
2. Streaming services integration: ListenBrainz integrates with popular streaming services like Spotify, enabling users to effortlessly link their accounts and contribute their streaming listening history.
3. Mobile applications and browser extensions: Various mobile applications can be used to submit listen events from mobile devices. Browser Extensions like WebScrobbler [10] provide convenient tools for submitting listening data from web-based music platforms.
4. Import of streaming services data exports: ListenBrainz supports conventional data donation methods, allowing users to upload data packages from streaming platforms like Spotify's extended streaming data export.

It is important to note that ListenBrainz empowers users with complete control over their data. They can edit, delete, or export their listening history as desired, ensuring transparency and user agency.

```
{
  "user_id": 1,
  "user_name": "rob",
  "timestamp": 1720644002,
  "track_metadata": {
    "track_name": "Tokara",
    "artist_name": "Fakear",
    "release_name": "All Glows",
    "additional_info": {
      "duration_ms": 206230,
      "tracknumber": 9,
      "artist_mbids": [
        "7c707d22-1c9c-4e72-bc8d-640baa5e2ba5"
      ],
      "release_mbid":
      ↪  "2524b5bd-03d2-48ea-b85c-8cdebc8bbfe4",
      "recording_mbid":
      ↪  "ba97f6e5-f4ff-404f-b95b-e3aabade5e2e",
      "submission_client": "navidrome",
      "submission_client_version": "0.51.0
      ↪  (fd61b29a)",
      "recording_msid":
      ↪  "886bf922-8041-4e02-9991-596ffebddb7a"
    }
  },
  "recording_msid":
  ↪  "886bf922-8041-4e02-9991-596ffebddb7a"
}
```

Listing 1: A listen event in the ListenBrainz dataset

4.2 Data Cleaning and Synthesis

ListenBrainz ensures data quality through a robust cleaning and synthesis process. Every listening event requires a UTC epoch timestamp, a user identifier assigned by ListenBrainz, track name, and artist name. The *additional_info* field permits users to submit free-form JSON data. This flexibility empowers users to contribute any relevant information they deem valuable, fostering a richer understanding of music listening behaviors. Commonly used additional metadata fields include release name, MusicBrainz identifiers, track position, duration, and music service or media player used. A MBID is a 36 character

[7] ListenBrainz API documentation is available at https://listenbrainz.readthedocs.io/en/latest/users/api-usage.html

[8] A list of known music player supporting ListenBrainz submission can be found at https://listenbrainz.org/add-data/

[9] Last.fm compatible API documentation at https://listenbrainz.readthedocs.io/en/latest/users/api-compat.html

[10] WebScrobbler https://web-scrobbler.com/

Universally Unique Identifier that is permanently assigned to each entity in the MusicBrainz database. The range of MusicBrainz identifiers (MBIDs) supported by the ListenBrainz dataset is broader than MLHD [3] and hence, opens doors to a wealth of additional information. For example, a release MBID allows access to detailed label data and cover art from the MusicBrainz ecosystem. Listing 1 shows an example of a listen history event in the ListenBrainz dataset.

To prevent duplicates, ListenBrainz employs a real-time deduplication system based on the unique combination of user ID, timestamp, and a MessyBrainz identifier (MSID). MSIDs are random UUIDs assigned to the hash of the track, artist, and release names, serving as a robust method for identifying unique listening events.

While submitting MusicBrainz identifiers (MBIDs) alongside listening events greatly enhances the dataset's connectivity and analytical potential, it's not always a straightforward task for users. Local music collections often lack MBIDs in their ID3 tags, necessitating additional efforts to improve metadata quality. ListenBrainz encourages users to utilize tools like MusicBrainz Picard [11] to tag their collections effectively. Tagging collections becomes impractical when users engage with music through streaming services, where control over metadata submission is limited. To address this challenge, ListenBrainz employs a sophisticated background service known as the MBID mapper. This service automatically searches and associates relevant MBIDs with listening events based on the available metadata, enriching the dataset's interconnectedness which is very helpful in downstream analysis. The inner workings of the MBID mapper involve complex algorithms and matching techniques beyond the scope of this paper. The MBIDs linked by the mapper are stored separately from user-submitted identifiers, empowering users of the dataset to choose whether or not to incorporate them into their analyses.

4.3 Dataset Format and Updates

The ListenBrainz dataset is available in two formats: ListenBrainz full export Dumps and ListenBrainz Spark Dumps. The ListenBrainz full export dumps contain the entire data submitted to ListenBrainz split in monthly chunks. Monthly data is organized into JSON lines files within yearly directories, providing comprehensive information for each listening event. The ListenBrainz spark dumps consist of chronologically ordered parquet files offering a subset of relevant fields optimized for batch processing and analysis.

The entire dataset is updated every 15 days, while incremental dumps capturing the listening events of the last 24 hours are produced daily. This ensures researchers and developers have access to both the comprehensive historical record and the most recent trends in music consumption.

[11] MusicBrainz Picard https://picard.musicbrainz.org/

5. DATASET ANALYSIS

As of today, the ListenBrainz listens dataset boasts a substantial collection of 876 million listening events contributed by approximately 28,000 users. Impressively, 764 million of these entries have been successfully linked with MusicBrainz identifiers, allowing for deeper analysis and connections with external music information resources. The dataset encompasses a diverse musical landscape, representing 900 thousand artists, 2.07 million albums, and a staggering 12.1 million recordings. Table 3 provides a summary of these key figures and a comparison with the corresponding figures of the MLHD dataset.

	MLHD [3]	**ListenBrainz**
Users	583 K	28,419
Listens (All)	27 B	876 M
Listens (with MBIDs)	-	764 M
Recordings	7 M	12.1 M
Albums	900 K	2.07 M
Artists	555 K	900 K

Table 3: Comparision of the size of the MLHD and ListenBrainz dataset

While the number of users and listening events in ListenBrainz is currently smaller compared to MLHD, it excels in its coverage of musical content, with several times the number of unique recordings, albums, and artists represented. This richness shows the potential of ListenBrainz for exploring a wider range of musical tastes and preferences.

The additional metadata recorded by ListenBrainz introduces several innovative features not present in the MLHD dataset. Specifically, 11% of listening events in ListenBrainz include track number information, while 12% of entries offer track duration data, which facilitates the analysis of listening session lengths and potential skipping behaviors. Additionally, 68% of listening events record the submission client. Although more than half of these clients are from Last.fm imports and Spotify, the remaining entries encompass a diverse array of user setups, including self-hosted music servers such as Navidrome and Funkwhale, as well as popular applications like Plex, PanoScrobbler, and WebScrobbler. This additional metadata enables new research opportunities to explore platform-specific listening behaviors and the influence of various music access modes on consumption patterns.

The temporal span of the ListenBrainz dataset is noteworthy, encompassing listening events dating back to 2005 and extending to the present year, 2024. Figure 2 illustrates the distribution of listening events across different years. Notably, the ability to submit past listening data to ListenBrainz suggests that the representation of earlier years may continue to grow over time. The lower number of events for 2024 is expected, given that only a portion of the year has elapsed.

Figure 1 shows the global coverage of the ListenBrainz

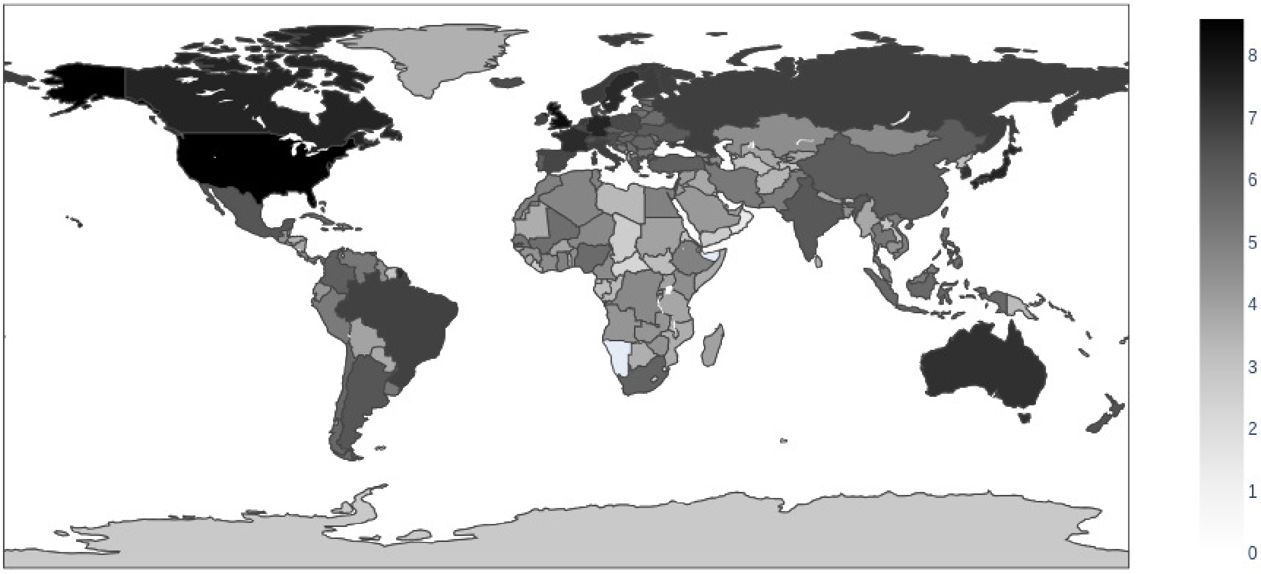

Figure 1: Artist Origins: Logarithm of number of listens of artists originating from a country

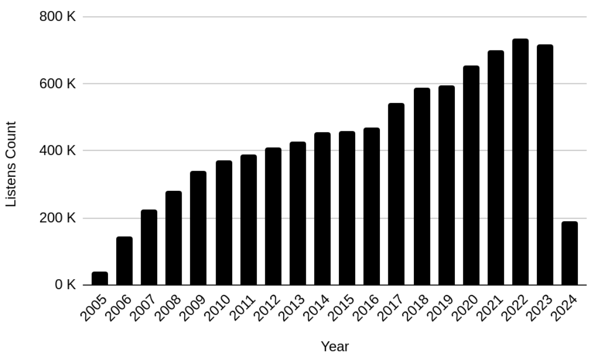

Figure 2: Temporal distribution of listening events

dataset, the artists in the dataset originate from a wide array of countries and regions. Although there is a noticeable concentration of artists originating from the United States, as evidenced by the darker shading, the dataset encompasses a diverse representation of artists from across the world particularly prominent in Europe, parts of South America, and Australia. This exploration also acts as an example of how MBIDs in the ListenBrainz dataset can be used to obtain useful information from the MusicBrainz database, in this case the country of an artist's origin.

Figure 3 displays another temporal aspect of the dataset, the distribution of listening events based on the release year of the music. The graph reveals a clear trend towards a preference for newer music, with a significant surge in listening events observed from the 1990s onwards. This pattern aligns with the increasing availability and accessibility of digital music during this period. Nevertheless, the presence of listening events for music spanning several decades, dating back to the 1960s and earlier, emphasizes the assorted range of musical interests within the ListenBrainz community and the enduring appeal of older music.

6. USE CASES

The dataset is actively by the ListenBrainz project itself internally to power collaborative filtering algorithms that generate personalized recommendations, playlists, and engaging user reports. By combining these collaborative filtering techniques with content-based recommendations derived from MusicBrainz's genre and folksonomy data, ListenBrainz creates a multifaceted and tailored music discovery experience for its users [12]. In a further commitment to open-source music recommendation development, the ListenBrainz team has created the Troi recommendation toolkit [13]. This standalone toolkit adopts an API-first philosophy, enabling the construction of diverse and engaging playlists by utilising ListenBrainz data alongside other compatible datasets. Similarly, the Calliope project is an external initiative that leverages the ListenBrainz dataset to curate playlists and aid research and development in the field of open-source music recommendation systems [14].

Beyond its applications in understanding general music preferences and trends, the listens data in ListenBrainz has proven valuable in exploring the impact of music recommendation diversity on listeners' long-term attitudes and engagement [14]. Researchers have leveraged listens data available in ListenBrainz to develop and evaluate sequential music recommendation systems that utilize the powerful BERT transformer model [15].

Like MLHD, ListenBrainz is built upon a foundation of user-generated listening histories, making it conceptually similar and offering comparable data for analysis. Although the user base and overall size differ, the core data structure allows for the application of similar research methodologies and comparisons between findings. Addi-

[12] Weekly Recommendation Playlists https://community.metabrainz.org/t/our-weekly-recommendations-are-now-live/646950?u=lucifer

[13] Troi recommendation toolkit https://troi.readthedocs.io/en/latest/

[14] Calliope Project https://calliope-music.readthedocs.io/

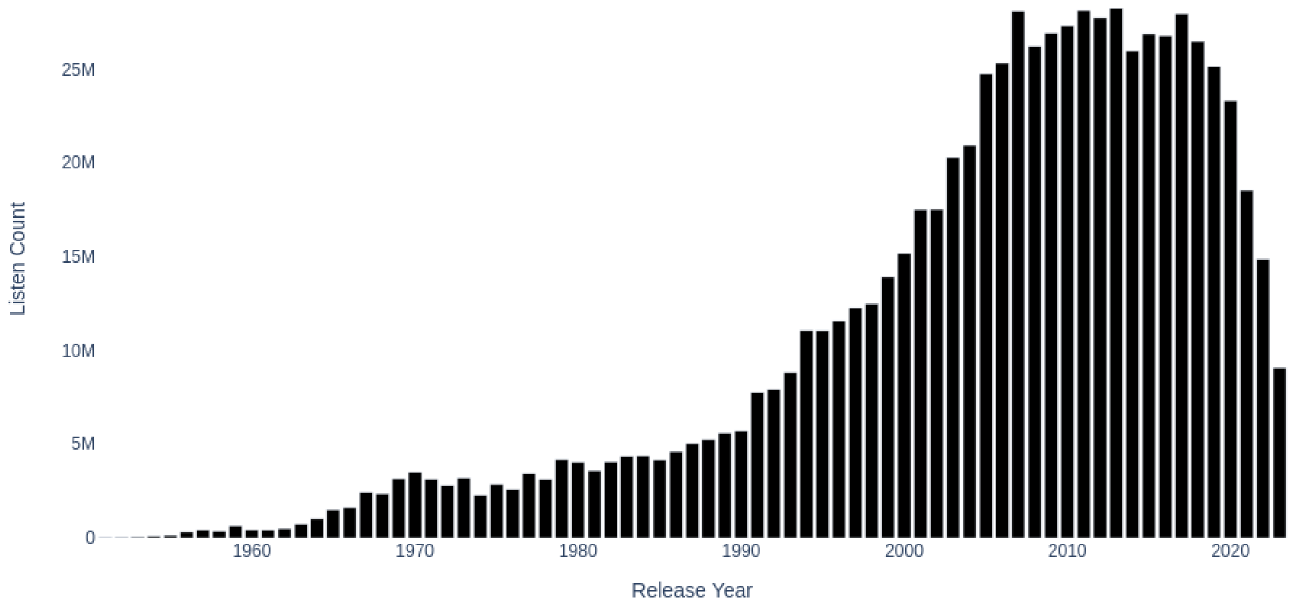

Figure 3: Listening events by original release year of albums

tionally, music sessions can be extracted from individual listening events of the ListenBrainz dataset to reproduce and extend studies initially conducted on the session-based MSSD dataset. Consequently, we are certain that this dataset holds immense potential for reproducing and validating previously conducted studies on similar datasets.

ListenBrainz also presents itself as a significant advancement in music consumption research tools. Instead of developing custom data collection and processing tools for data donations, researchers can leverage ListenBrainz. Researchers are relieved from the technical burdens and logistical complexities of data collection, allowing them to dedicate their time and resources to the core aspects of their studies and analytical inquiries. The ListenBrainz platform provides participants with insights into their listening behavior which can potentially increasing study engagement as well. In return, the listening events submitted by the participants enrich the overall listens dataset.

7. LIMITATIONS

The dataset utilizes UTC timestamps which prevents its usage in temporal analyses involving time zones, such as the diurnal music preferences explored by Park et. al [10]. Future iterations of the dataset aim to incorporate timestamps aligned with users' respective time zones, further enhancing its analytical capabilities.

The dataset can only as diverse as the individuals who choose to share their listening histories, potentially creating limitations in representing the full spectrum of music consumption across various cultures, genres, and communities. For instance, Figure 1 reveals a geographic bias in ListenBrainz's user demographics, with a disproportionate number of users located in the Anglosphere. Efforts are underway to integrate demographic data, such as user region and gender, to provide additional context to detect and eliminate such biases.

An inherent challenge within music listening datasets, including ListenBrainz, is the difficulty in discerning whether a listening event reflects a user's genuine music preference or merely their exposure to a track due to algorithmic recommendations or shuffle mechanisms within music streaming services. This ambiguity makes it difficult to determine if a specific listening event represents an active choice by the user or a passive encounter with a suggested track.

Further, growing concerns surrounding online privacy may lead users to be hesitant in sharing their personal data, including seemingly benign information like music listening habits, impacting the growth of the dataset. Individuals are becoming increasingly aware of data collection practices and harbor reservations about potential privacy risks and the possible misuse of their information [16].

8. CONCLUSION

In conclusion, the ListenBrainz listens dataset provides a rich and dynamic resource for understanding the complexities of music consumption. Its comprehensive collection of user listening histories, accurate to the second, offers valuable insights into individual preferences and general trends. The inclusion of MusicBrainz identifiers further enhances its utility, enabling seamless integration with external music databases and facilitating in-depth analyses.

To reiterate, the ListenBrainz listens dataset addresses a significant gap in the field by providing a continuously updated resource that can represent rapidly changing music preferences. As the ListenBrainz project is run by a non-profit entity devoid of vested corporate interests, we believe that it will emerge as an indispensable resource for future research endeavors. By embracing openness, user agency, and continuous growth, ListenBrainz listens dataset paves the way for a deeper understanding of how we engage with music.

The dataset can be downloaded from `https://data.metabrainz.org/pub/musicbrainz/listenbrainz/fullexport/`.

9. REFERENCES

[1] S. M. West, "Data capitalism: Redefining the logics of surveillance and privacy," *Business & Society*, vol. 58, no. 1, pp. 20–41, 2019. [Online]. Available: https://doi.org/10.1177/0007650317718185

[2] K. Choi, G. Fazekas, K. Cho, and M. Sandler, "A tutorial on deep learning for music information retrieval," 2018.

[3] G. Vigliensoni and I. Fujinaga, "The music listening histories dataset," in *Proceedings of the 18th International Society for Music Information Retrieval Conference*, Suzhou, People's Republic of China, 2017, pp. 96–102.

[4] B. Brost, R. Mehrotra, and T. Jehan, "The music streaming sessions dataset," in *The World Wide Web Conference*, ser. WWW '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 2594–2600. [Online]. Available: https://doi.org/10.1145/3308558.3313641

[5] O. Celma, *Music recommendation and discovery*. Springer-Verlag Berlin Heidelberg, 2010.

[6] M. Schedl, "The lfm-1b dataset for music retrieval and recommendation," in *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, ser. ICMR '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 103–110. [Online]. Available: https://doi.org/10.1145/2911996.2912004

[7] A. B. Melchiorre, N. Rekabsaz, E. Parada-Cabaleiro, S. Brandl, O. Lesota, and M. Schedl, "Investigating gender fairness of recommendation algorithms in the music domain," *Information Processing & Management*, vol. 58, no. 5, p. 102666, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306457321001540

[8] G. Vigliensoni and I. Fujinaga, "Identifying time zones in a large dataset of music listening logs," in *Proceedings of the First International Workshop on Social Media Retrieval and Analysis*, ser. SoMeRA '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 27–32. [Online]. Available: https://doi.org/10.1145/2632188.2632203

[9] P. Herrera, Z. Resa, and M. Sordo, "Rocking around the clock eight days a week: an exploration of temporal patterns of music listening," in *1st Workshop On Music Recommendation And Discovery (WOMRAD), ACM RecSys, 2010, Barcelona, Spain*, 2010.

[10] M. Park, J. Thom, S. Mennicken, H. Cramer, and M. Macy, "Global music streaming data reveal diurnal and seasonal patterns of affective preference," *Nature Human Behaviour*, vol. 3, no. 3, pp. 230–236, Mar 2019. [Online]. Available: https://doi.org/10.1038/s41562-018-0508-z

[11] J. Ohme, T. Araujo, L. Boeschoten, D. Freelon, N. Ram, B. B. Reeves, and T. N. Robinson, "Digital trace data collection for social media effects research: Apis, data donation, and (screen) tracking," *Communication Methods and Measures*, vol. 18, no. 2, pp. 124–141, 2024. [Online]. Available: https://doi.org/10.1080/19312458.2023.2181319

[12] N. Pfiffner and T. N. Friemel, "Leveraging data donations for communication research: Exploring drivers behind the willingness to donate," *Communication Methods and Measures*, vol. 17, no. 3, pp. 227–249, 2023. [Online]. Available: https://doi.org/10.1080/19312458.2023.2176474

[13] G. Mazziotti and H. Ranaivoson, "Can online music platforms be fair? an interdisciplinary research manifesto," *IIC - International Review of Intellectual Property and Competition Law*, vol. 55, no. 2, pp. 249–279, Feb 2024. [Online]. Available: https://doi.org/10.1007/s40319-023-01420-w

[14] L. Porcaro, E. Gómez, and C. Castillo, "Assessing the impact of music recommendation diversity on listeners: A longitudinal study," *ACM Trans. Recomm. Syst.*, vol. 2, no. 1, mar 2024. [Online]. Available: https://doi.org/10.1145/3608487

[15] N. Yadav and A. K. Singh, "Bi-directional encoder representation of transformer model for sequential music recommender system," in *Proceedings of the 12th Annual Meeting of the Forum for Information Retrieval Evaluation*, ser. FIRE '20. New York, NY, USA: Association for Computing Machinery, 2021, p. 49–53. [Online]. Available: https://doi.org/10.1145/3441501.3441503

[16] A. Acquisti, L. Brandimarte, and G. Loewenstein, "Privacy and human behavior in the age of information," *Science*, vol. 347, no. 6221, pp. 509–514, Jan. 2015.