# LFM-2b: A Dataset of Enriched Music Listening Events for Recommender Systems Research and Fairness Analysis

Markus Schedl
Stefan Brandl
Oleg Lesota
Emilia Parada-Cabaleiro
David Penz
Navid Rekabsaz
markus.schedl@jku.at
Johannes Kepler University (JKU) and
Linz Institute of Technology (LIT)
Linz, Austria

## ABSTRACT

We present the LFM-2b dataset containing the listening records of over 120,000 users of the music platform Last.fm. These users provide a total of more than two billion individual listening events that span a time range of over 15 years, from February 2005 until March 2020. These listening events refer to a total of 50 million distinct tracks of 5 million distinct artists. Beside the common metadata (i. e., artist and track name), LFM-2b contains additional information both regarding the users and items. This includes the demographic information of users, namely country, gender, and age, and the fine-grained genre and style of items together with the vector embeddings of their lyrics.

LFM-2b is a rich dataset that enables research on a variety of recommender system algorithms, such as the ones based on collaborative filtering (e.g., leveraging the user–item interactions in the form of listening events), but also content-based approaches (e.g., exploiting genres and lyrics), or hybrid combinations thereof. Users' demographic information furthermore enable experimentation on identifying and mitigating various data and algorithmic biases of recommender systems, and investigating fairness aspects of such systems, e.g., according to gender.

## KEYWORDS

recommender systems, user modeling, music information retrieval, dataset, experimentation, bias, fairness, classification, auto-tagging

**ACM Reference Format:**
Markus Schedl, Stefan Brandl, Oleg Lesota, Emilia Parada-Cabaleiro, David Penz, and Navid Rekabsaz. 2022. LFM-2b: A Dataset of Enriched Music Listening Events for Recommender Systems Research and Fairness Analysis. In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '22), March 14–18, 2022, Regensburg, Germany.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3498366.3505791


*CHIIR '22, March 14–18, 2022, Regensburg, Germany*

ACM ISBN 978-1-4503-9186-3/22/03.
https://doi.org/10.1145/3498366.3505791

## 1 MOTIVATION

Recommender systems (RSs) have become highly influential automatic decision making tools that strongly affect our daily lives. While being adopted in a variety of domains, such as e-commerce, news, travel, and recruiting, the entertainment domain, in particular movies and music, has historically been at the forefront of research interest in RSs.

Researching, developing, and especially appropriately evaluating novel RS algorithms requires publicly available datasets of user–item interactions and item content descriptors, for collaborative filtering (CF) and content-based filtering (CBF) approaches, respectively. Such datasets should ideally include data created by real users in the wild, to provide a realistic experimental playground.

The LFM-2b dataset is such a stable and standardized repository. It provides interaction, content, and contextual information, which facilitates reproducible experimentation in the music recommendation domain for a range of algorithms such as CF, CBF, context-aware, and hybrid recommenders. The unique composition of LFM-2b (including user demographics and item content descriptors) furthermore enables its use beyond traditional music RS tasks, for instance for various data science experiments [7], intervention time series analysis [13], uncovering various biases in real-world data and recommendation algorithms [8], and evaluating bias mitigation strategies for improved recommendation fairness [10].

LFM-2b is based upon the dataset we released as part of [10], which is itself an extension of our earlier LFM-1b dataset. In addition to these earlier versions, the current and stable version of the dataset adds (1) user-generated tags on the track level, (2) fine-grained genre labels on the track level, (3) text embeddings of lyrics, and (4) Spotify identifiers to facilitate an easy extension (e.g., by Spotify's audio features).

## 2 RELATED DATASETS

There exist several datasets suited to build and evaluate music RSs. Such datasets highly differ in terms of their sizes, ages, the variety and coverage of the data they contain, as well as the data sources their creators leveraged to obtain listening records and side information. The most prominent datasets include the Million Song Dataset [1], Spotify's Music Streaming Sessions Dataset [2]

and Million Playlist Dataset [16], the AotM-2011 dataset of playlists from Art of the Mix[1] [9], and the Yahoo! Music Dataset [5]. Some datasets are built from listening information extracted from Twitter,[2] for instance the Million Musical Tweets Dataset [6] and the #nowplaying-RS dataset [11].

In addition to the mentioned ones, a few datasets leverage Last.fm data. Such datasets include earlier (and meanwhile outdated) Last.fm 360K and Last.fm 1K [3], as well as the more recent and considerably sized Music Listening Histories Dataset [14], and LFM-1b [12], which is in fact the predecessor of LFM-2b, our dataset provided in the work at hand.

In contrast to the mentioned datasets, the proposed LFM-2b dataset provides a unique resource with the following features:

- *Large scale:* LFM-2b contains more than 2 billion listening events created by more than 120 thousand users.
- *Wide timeframe:* LFM-2b provides listening data that extends from 2005 to 2020, therefore covering 15 years.
- *Usage variety:* LFM-2b contains a unique combination of collaborative data (user–item interactions), content data (tags and lyrics embeddings), and contextual data (user characteristics)
- *Enriched by demographic information:* The data of LFM-2b users includes age, gender, and country information, which enables the study of population biases in data and RS algorithms, among other possible use cases.

## 3 THE LFM-2B DATASET

LFM-2b is a dataset of music listening events (LEs) created by users of the music platform Last.fm.[3] Last.fm empowers its users to store their listening records across all their devices in a central location, and to share them with others. It also integrates a music recommender system. A LE represents an interaction between a user and a music track (i.e., a user listening to a track), enriched by metadata of the interaction. More specifically, a LE is defined as <user-id, artist-id, track-id, album-id, timestamp>. The timestamp refers to the starting time of the LE, provided at the granularity of seconds. Note that privacy of users is preserved by only storing an incremental numeric user identified that cannot be traced back to a particular Last.fm user name.

This data can already be leveraged for studying recommender systems based on *collaborative filtering*. LFM-2b furthermore includes additional item-level information, namely user-generated tags, fine-grained genre labels, and embedding vectors of lyrics. This data enables the study and development of *content-based* recommender engines.

In addition, the data of some of the (anonymized) users are enriched with their demographic information (country, age, and gender). This enables research on detailed user modeling for improved personalization of recommendations, as well as investigation on *biases* in data and *fairness* of recommender algorithms.

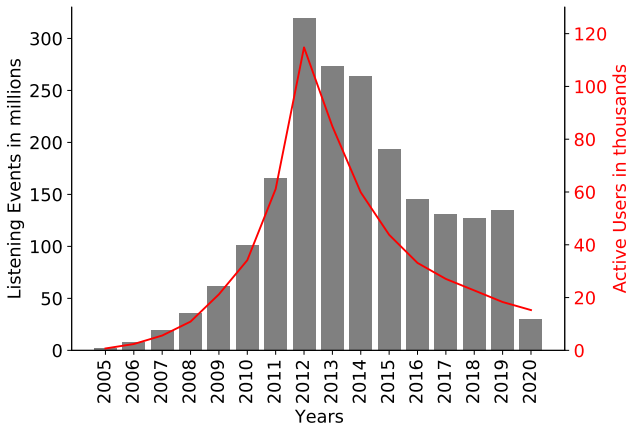


**Figure 1: Listening events and active users by year**

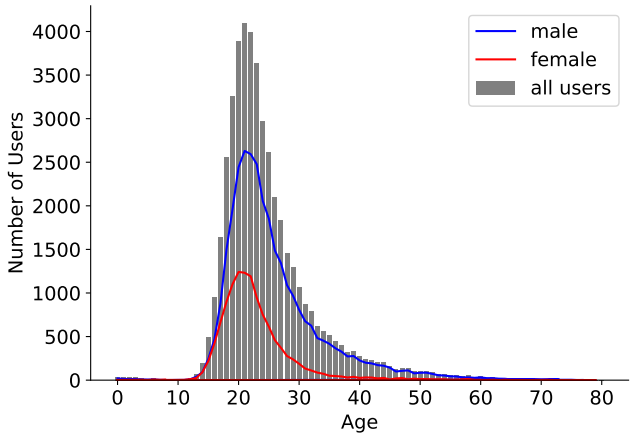


**Figure 2: Histograms of all/female/male users by age**

The dataset, accompanied with the Python code to load and use the data, are publicly available at http://www.cp.jku.at/datasets/LFM-2b.[4]

### 3.1 Dataset Content and Statistics

The LEs of the LFM-2b dataset span a time range from February 14, 2005 until March 20, 2020, making a total of 2,014,164,872 LEs. About half of these LEs were created between 2005 and the end of 2014 (9 years), and the other half in the subsequent 5 years, namely between 2015 to 2020. Figure 1 reports the statistics of the LEs as well as the number of active users per year. As shown, starting from 2012, active users are decreasing over years. This leads to a decrease in the number of LEs from 2012, while the decrease saturates after 2016, indicating an increase in the per-user consumption in this time period.

In the following, we explain in detail the various data elements provided in the LFM-2b dataset. The statistics and the data format

[1] http://www.artofthemix.org
[2] https://www.twitter.com
[3] https://www.last.fm

[4] The LFM-2b dataset is considered derivative work according to §4.1 of Last.fm's API Terms of Service (https://www.last.fm/api/tos). The Last.fm Terms of Service further grant us a license to use and publish this data (according to §4).

| Dataset | No. | Columns |
|---|---|---|
| albums | 24,237,348 | album_id, album_name, artist_name |
| artists | 5,159,580 | artist_id, artist_name |
| listening-events | 2,014,164,872 | user_id, track_id, album_id, timestamp |
| spotify-uris | 4,624,359 | track_id, uri |
| tracks | 50,813,373 | track_id, artist_name, track_name |
| listening-counts | 519,293,333 | user_id, track_id, count |
| users | 120,322 | user_id, country, age, gender, creation_time |
| lyrics-features | 1,266,554 | track_id, features{...} |
| tags | 2,230,814 | track_id, <tag, weight>+ |
| tags-micro-genres | 1,638,468 | track_id, <micro-genre, weight>+ |

**Table 1: Statistics and the format of the columns of the data available in the LFM-2b dataset**

| | Total no. | Min | Q1 | Median | Q3 | Max | Mean | Std |
|---|---|---|---|---|---|---|---|---|
| Artists per user | 5,159,580 | 1.0 | 162.0 | 389.0 | 984.0 | 169,202.0 | 881.0 | 1,565.78 |
| Tracks per user | 50,813,373 | 1.0 | 525.0 | 1,596.0 | 5,159.0 | 243,372.0 | 4,315.86 | 7,464.68 |
| LEs per user | 2,014,164,872 | 1.0 | 1,065.0 | 4,147.5 | 19,921.0 | 946,750.0 | 16,739.79 | 33,721.58 |
| Tags per track | 2,230,814 | 1.0 | 2.0 | 5.0 | 11.0 | 100.0 | 11.37 | 18.73 |
| Micro-genres per track | 1,638,468 | 1.0 | 1.0 | 2.0 | 4.0 | 60.0 | 3.04 | 3.12 |

**Table 2: LFM-2b's statistics based on user- and track-based metrics**

in the corresponding data sheet regarding each of the data elements are summarized in Table 1.

- artists: name of 5,159,580 artists.
- albums: name of 24,237,348 albums, accompanied with the names of their artists.
- tracks: name of 50,813,373 tracks, accompanied with their artists.
- users: information of 120,322 users, containing country, age, gender, and creation-time. Country is specified according to ISO 3166 Alpha-2 country code; empty if unknown. Age is the age of the user; −1 if unknown. Gender is either "m" (male), "f" (female), or "n" (neutral); empty when no gender information is present. Creation-time indicates the time that the user profile is created. Figure 2 reports the histogram of users over age, separately shown for all, female, and male users.
- listening-events: 2,014,164,872 LEs, where each data point consists of the ID of the user, the ID of the track and the album, and the timestamp of the event. Artist can be inferred from tracks using column track_id.
- listening-counts: has 519,293,333 records, containing the number of times a user has listened to a certain track.
- spotify-uris: the URI of 4,624,359 tracks is provided, which can be used for crawling audio features or additional metadata from Spotify. Note that URIs are only specified for the tracks in the LFM-2b dataset which are also included in Spotify's catalog.
- lyrics-features: provides 1,266,554 records, containing the lexical features, compression ratios, entropy values, and vector embeddings of the lyrics of the subset of tracks for which we could retrieve lyrics.
- tags: for a subset of 2,230,814 tracks, the user-generated tags are provided. Each of these tracks are annotated by users with one or more tags in the form of <tag, weight> pairs (tags=<tag, weight>+). Weights are values between 1 and 100 rounded to the nearest integer. The tag with the most annotations for a given song gets a weight of 100, and all other weights are set to the relative percentages of the most common one. Overall, there are 1,041,819 unique tags in the dataset.
- tags-micro-genres: we also provide a subset of tags, containing 1,638,468 records exclusively with the information of micro-genres, i. e., fine-grained indications of musical genres or styles. The process of extracting micro-genres is explained in Section 3.2. The top 10 of the micro-genres and their statistics are reported in Table 3.

Table 2 reports the seven-number statistical summary of the various user- and track-based proportions in LFM-2b. The explained elements of LFM-2b are provided in the tab-separated files (tsv) for tabular data, and in JSON format for large and bulky textual data.

## 3.2 Data Acquisition and Processing

Adapting the methodology of acquiring the LFM-1b dataset [12], LEs and user demographics of LFM-2b are collected from Last.fm, using the provided API.[5] Listenings events are retrieved from the user.getRecentTracks[6] endpoint, for the same users as contained in LFM-1b. The API response contains a timestamp for each listening event as well as a URL including artist name, album name

[5]https://www.last.fm/api
[6]https://www.last.fm/api/show/user.getRecentTracks

| Micro-genre | Abs. frequency | Rel. frequency |
|---|---|---|
| rock | 318,845 | 19.46% |
| pop | 196,888 | 12.02% |
| metal | 124,624 | 7.61% |
| alternative rock | 101,530 | 6.20% |
| jazz | 97,967 | 5.98% |
| ambient | 91,131 | 5.56% |
| folk | 81,962 | 5.00% |
| experimental | 80,558 | 4.92% |
| singer-songwriter | 76,634 | 4.68% |
| electronica | 76,352 | 4.66% |

**Table 3: Absolute and relative frequencies of the top 10 micro-genres**

(optionally), and track name. We extract unique tracks by grouping on the basis of <artist-name, track-name> tuple, since the album-name is optional. The same song can appear in multiple albums. Demographic information for each user are gathered via the `user.getInfo`[7] endpoint.

Tag information is retrieved for every track that is listened to at least 10 times via `track.getTopTags`[8] endpoint. We also compute a cleaned but still fine-grained list of tags by indexing tags using the micro-genres provided by Every Noise at Once.[9] We hence remove every tag not contained in this list of micro-genres.

In order to create lyrics features, we first crawl the tracks' lyrics from Genius[10] using the unique <artist-name, track-name> tuples. We process the text of the lyrics by removing special tokens such as "[Chorus]" or "[Intro]". The processed text of each track is then used to calculate the set of text features, namely: the compression-rate using `zlib`,[11] Shannon entropy, token-count, line-count, character-count, stop-word-count, and hyphen-count, as well as the lyrics embedding using the BERT-Base language model [4].[12]

To enhance the dataset with Spotify URIs,[13] we submit queries to Spotify's `tracks/id`[14] endpoint using the <artist-name, track-name> tuples. These URIs can be used, among others, to gather audio features via the `audio-features`[15] endpoint.

# 4 USE CASES

## 4.1 Music Recommendation

Besides traditional recommendation approaches such as collaborative filtering and content-based filtering, the `LFM-2b` dataset, containing user-generated tags, is additionally suitable to develop and evaluate context-aware recommendation algorithms. This could be realized, for instance, by taking into account the specific situations or moods with which musical items have been annotated, and which can be extracted from the provided tags.

## 4.2 Music Classification and Tagging

The `LFM-2b` dataset, containing information on musical genres and sub-genres as well as pre-computed features from the lyrics, enables the investigation of relationships between lyrics and musical genres, by this promoting tasks such as genre or mood classification. Similarly, since `LFM-2b` includes a mapping from tracks to Spotify identifiers (which enables retrieval of acoustic features), it is also suitable to carry out auto-tagging tasks, by leveraging audio descriptors as input features and mood labels extracted from user-generated tags as target classes.

## 4.3 Uncovering Biases and Unfairness

As mentioned before, a subset of users in `LFM-2b` are accompanied with their contextual and demographic information, namely their country, age, and gender metadata. This additional information provides a unique opportunity to study the bias and fairness in RSs, i. e., the discrepancies regarding the different treatment of users that belong to different demographic groups. This topic can be studied from the perspective of CF (leveraging user–item interactions), but also considering CBF recommendation, and hybrid approaches. For instance, a possible task in this context is the study of the (cor)relation of specific words or phrases in tags or lyrics with specific genders or age groups.

# 5 CONCLUSION AND POSSIBLE EXTENSIONS

We presented `LFM-2b`, a large and publicly available dataset of music listening events created by Last.fm users, enriched with demographic information and item content descriptors (tags, fine-grained genres, and lyrics embeddings). We detailed the characteristics of the dataset and described several use cases, including music recommendation, classification, and tagging, as well as identifying and mitigating various biases.

We envision several extensions to `LFM-2b`. For instance, we are currently working on enhancing the temporal coverage of listening events in `LFM-2b`, so that it spans an even longer time period including the peaks of the Covid-19 pandemic. Also, `LFM-2b` could be extended by emotion annotations extracted via textual emotion recognition techniques from tags or lyrics.

# ACKNOWLEDGMENTS

This research received support by the Austrian Science Fund (FWF): P33526.

# REFERENCES

[1] Thierry Bertin-Mahieux, Daniel P. W. Ellis, Brian Whitman, and Paul Lamere. 2011. The Million Song Dataset. In *Proceedings of the 12th International Society for Music Information Retrieval Conference, ISMIR 2011, Miami, Florida, USA, October 24-28, 2011*, Anssi Klapuri and Colby Leider (Eds.). University of Miami, 591–596. http://ismir2011.ismir.net/papers/OS6-1.pdf

[2] Brian Brost, Rishabh Mehrotra, and Tristan Jehan. 2019. The Music Streaming Sessions Dataset. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, Ling Liu, Ryen W. White, Amin Mantrach, Fabrizio Silvestri, Julian J. McAuley, Ricardo Baeza-Yates, and Leila Zia (Eds.). ACM, 2594–2600. https://doi.org/10.1145/3308558.3313641

[7] https://www.last.fm/api/show/user.getInfo
[8] https://www.last.fm/api/show/track.getTopTags
[9] https://everynoise.com/everynoise1d.cgi?scope=all (snapshot 2021-10-25)
[10] https://genius.com
[11] https://www.zlib.net
[12] Lyrics embeddings are computed using the `bert-base-uncased` model provided by the Huggingface [15] library
[13] https://community.spotify.com/t5/FAQs/What-s-a-Spotify-URI/ta-p/919201
[14] https://developer.spotify.com/documentation/web-api/reference/#/operations/search
[15] https://developer.spotify.com/documentation/web-api/reference/#/operations/get-several-audio-features

[3] Òscar Celma. 2010. *Music Recommendation and Discovery - The Long Tail, Long Fail, and Long Play in the Digital Music Space*. Springer, Berlin, Germany. https://doi.org/10.1007/978-3-642-13287-2

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 4171–4186.

[5] Gideon Dror, Noam Koenigstein, Yehuda Koren, and Markus Weimer. 2011. The Yahoo! Music Dataset and KDD-Cup'11. In *Proceedings of the 2011 International Conference on KDD Cup 2011 - Volume 18 (KDDCUP'11)*. JMLR.org, 3–18.

[6] David Hauger, Markus Schedl, Andrej Košir, and Marko Tkalčič. 2013. The million musical tweet dataset: what we can learn from microblogs. In *Proceedings of the International Society for Music Information Retrieval Conference*. Curitiva, Brazil, 189–194.

[7] Dominik Kowald, Peter Muellner, Eva Zangerle, Christine Bauer, Markus Schedl, and Elisabeth Lex. 2021. Support the underground: characteristics of beyond-mainstream music listeners. *EPJ Data Science* 10, 1 (2021), 1–26.

[8] Oleg Lesota, Alessandro B. Melchiorre, Navid Rekabsaz, Stefan Brandl, Dominik Kowald, Elisabeth Lex, and Markus Schedl. 2021. Analyzing Item Popularity Bias of Music Recommender Systems: Are Different Genders Equally Affected?. In *RecSys '21: Fifteenth ACM Conference on Recommender Systems, Amsterdam, The Netherlands, 27 September 2021 - 1 October 2021*, Humberto Jesús Corona Pampín, Martha A. Larson, Martijn C. Willemsen, Joseph A. Konstan, Julian J. McAuley, Jean Garcia-Gathright, Bouke Huurnink, and Even Oldridge (Eds.). ACM, 601–606. https://doi.org/10.1145/3460231.3478843

[9] Brian McFee and Gert RG Lanckriet. 2012. Hypergraph Models of Playlist Dialects. In *Proceedings of the International Society for Music Information Retrieval Conference*. ISMIR, Porto, Portugal, 343–348.

[10] Alessandro B. Melchiorre, Navid Rekabsaz, Emilia Parada-Cabaleiro, Stefan Brandl, Oleg Lesota, and Markus Schedl. 2021. Investigating gender fairness of recommendation algorithms in the music domain. *Inf. Process. Manag.* 58, 5 (2021), 102666. https://doi.org/10.1016/j.ipm.2021.102666

[11] A Poddar, E Zangerle, and Y Yang. 2018. nowplaying-RS: a new benchmark dataset for building context-aware music recommender systems. In *Proceedings of the 15th Sound and Music Computing Conference*.

[12] Markus Schedl. 2016. The LFM-1b Dataset for Music Retrieval and Recommendation. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval* (New York, New York, USA) *(ICMR '16)*. Association for Computing Machinery, New York, NY, USA, 103–110. https://doi.org/10.1145/2911996.2912004

[13] Markus Schedl, Eelco Wiechert, and Christine Bauer. 2018. The Effects of Real-world Events on Music Listening Behavior: An Intervention Time Series Analysis. In *Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon , France, April 23-27, 2018*, Pierre-Antoine Champin, Fabien Gandon, Mounia Lalmas, and Panagiotis G. Ipeirotis (Eds.). ACM, 75–76. https://doi.org/10.1145/3184558.3186936

[14] Gabriel Vigliensoni and Ichiro Fujinaga. 2017. The Music Listening Histories Dataset. In *Proceedings of the International Society for Music Information Retrieval Conference*. ISMIR, Suzhou, China, 96–102.

[15] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. 2019. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771* (2019).

[16] Hamed Zamani, Markus Schedl, Paul Lamere, and Ching-Wei Chen. 2019. An Analysis of Approaches Taken in the ACM RecSys Challenge 2018 for Automatic Music Playlist Continuation. *ACM Trans. Intell. Syst. Technol.* 10, 5 (2019), 57:1–57:21. https://doi.org/10.1145/3344257