Mood of the World: Analyzing Global and Seasonal Emotional Trends through Spotify Streaming Data

Jaime Gallego

Technische Universität Wien, Vienna, Austria
jaime.gallego@student.tuwien.ac.at

Abstract. Social media and streaming platforms serve as digital mirrors of collective human behavior. This paper presents "Mood of the World," a study analyzing the global emotional state through popular music using Spotify data. We aim to determine if cultural or seasonal patterns influence the "musical mood" of a nation. Facing recent restrictions in the Spotify Web API regarding audio features, we implemented a hybrid data collection strategy combining public datasets with API-fetched metadata and daily streaming charts. We define a "Mood Index" based on audio features (valence and energy) and introduce a "Weighted Mood Index" to account for streaming volume. The analysis covers top tracks from 8 countries across Europe, the Americas, Asia, and Oceania during opposing seasonal periods (August 2017 vs. January 2018). Results reveal distinct cultural baselines and noticeable seasonal fluctuations, particularly in European countries, validating the potential of streaming data as a social sensor.

Reproducibility: The source code and data processing scripts required to reproduce the results are available at:
https://github.com/jaimegallegoo/mood-of-the-world

Keywords: Social Media Analytics · Spotify API · Music Information Retrieval · Seasonal Trends · Mood Analysis.

1 Introduction

1.1 Motivation

Social media and digital streaming platforms have evolved beyond simple content delivery systems; they have become real-time sensors of collective human behavior. Every play, skip, and playlist addition contributes to a massive dataset that reflects not only individual preferences but also broader cultural and emotional trends.

The global success of data-driven campaigns like *Spotify Wrapped* or analytics platforms such as *stats.fm* demonstrates a growing public appetite for understanding personal behavior through streaming data. However, while these initiatives focus on the *individual*, this project aims to shift the lens to the *collective*.

Music, in particular, is intrinsically linked to emotion. Research suggests a strong tendency toward *mood congruency*, where listeners unconsciously select tracks that resonate with their current affective state, often gravitating towards upbeat, high-energy rhythms during moments of joy, or seeking solace in slower, melancholic melodies when feeling down. Unlike text-based social media (e.g., Twitter/X), where sentiment analysis requires complex natural language processing, music possesses objective acoustic characteristics, such as tempo, energy, and valence (a measure of musical positiveness), that can quantify the "mood" of a track directly.

This project, titled "Mood of the World," stems from a central hypothesis: music consumption is not merely a matter of personal taste, but is significantly influenced by external collective factors such as geography (culture) and seasonality (weather and light cycles). Does a country "sound" happier in summer than in winter? Is there a quantifiable "emotional gap" between the music consumed in Latin America versus Northern Europe?

1.2 Objectives

The primary objective of this study is to quantify and visualize the "musical mood" of different nations. To achieve this, we aim to:

- Define a robust metric, the *Mood Index*, based on Spotify's psycho-acoustic attributes (*valence* and *energy*).
- Analyze the top-tier music charts of 8 strategic countries across three continents (Europe, Americas, Asia/Oceania) to identify cultural baselines.
- Compare the musical mood across opposing seasons (August vs. January) to determine if seasonal changes correlate with significant shifts in the emotional tone of popular music.

1.3 Contribution

This paper makes two distinct contributions to the field of Social Media Analytics. First, on a technical level, we address the growing challenge of "API decay." Following recent restrictions by Spotify that blocked access to critical audio features for developers in development mode (returning HTTP 403 errors), we propose and validate a **hybrid data collection framework**. This framework successfully reconstructs the necessary dataset by fusing live API metadata, static public repositories, and historical streaming charts, utilizing a custom entity resolution pipeline to link them.

Second, on an analytical level, we introduce the **Weighted Mood Index**. Unlike previous studies that might average the features of a "Top 50" list equally, our metric weights each song's emotional score by its actual streaming volume. This approach ensures that a viral hit with millions of plays has a proportionally greater impact on the country's mood score than a song at the bottom of the chart, providing a more accurate representation of collective listening behavior.

2 Related Work

This project connects two main fields: Social Media Analytics and Music Information Retrieval (MIR). The goal is to move beyond analyzing text comments and instead look at music consumption to understand how groups of people feel.

2.1 Social Media as a Digital Sensor

In the past, researchers studied collective behavior using small surveys or direct observation. These methods were slow and often inaccurate because people act differently when they know they are being watched. Today, digital platforms act as real-time "social sensors." As Shirky notes [1], social media has changed how groups communicate and organize, creating huge amounts of data about human behavior.

Fan and Gordon [2] argue that analyzing this data helps us find patterns in public sentiment that were previously invisible. However, there is a big difference between "stated preferences" (what users say they like in public posts) and "revealed preferences" (what users actually consume in private). Analyzing text from Twitter or Facebook is difficult because users often try to project a specific image of themselves. In contrast, streaming data offers a more honest and direct signal of emotional states because it reflects what people actually choose to hear.

2.2 Music Information Retrieval and Emotions

Music is closely linked to how we regulate our emotions. In the field of Music Information Retrieval (MIR), many studies have tried to map audio features to human feelings. Hu and Downie [3] showed that we can predict mood effectively by combining audio data with metadata like genre or artist.

This project uses the "circumplex model of affect" proposed by Russell. This theory simplifies emotions by placing them on a graph with two axes:

1. **Valence:** This measures positivity. It ranges from unpleasant (sad or stressed) to pleasant (happy or satisfied).
2. **Arousal:** This measures intensity. It ranges from inactive (calm or bored) to active (excited or tense).

Fortunately, Spotify provides two audio features that match this model very well: `valence` and `energy`. By using these values, we can calculate a "Mood Index" mathematically without needing to interpret song lyrics.

2.3 Global Algorithms vs. Local Culture

Finally, we must consider the role of the platform itself. Streaming services are not neutral; they actively shape our taste. Recent studies suggest that recommendation algorithms tend to standardize listening habits around the world [7]. These systems often prioritize songs that keep users engaged, which creates a loop where the same global hits become popular everywhere.

This "algorithmic standardization" might be erasing local musical differences. However, it also gives us a common standard to compare countries. This project aims to test if local factors, such as the weather or the season, still influence what people listen to. For example, we want to see if the "Winter Blues" still exists in Europe despite the algorithm pushing global pop hits. By comparing charts from different seasons, we can see if the environment still affects the collective mood.

3 Methodology

3.1 Data Sources: The Hybrid Approach

The primary challenge in analyzing large-scale music emotion data is the accessibility of audio features. While the Spotify Web API historically provided endpoints to retrieve metrics such as *valence* and *energy* for any track, recent changes to the developer policy have restricted access to the `Get Tracks' Audio Features` endpoint for non-commercial applications in "Development Mode" returning a 403 Forbidden error.

To overcome this limitation without sacrificing the granularity of the analysis, we designed a hybrid data collection strategy that integrates three distinct sources:

1. **Historical Streaming Charts:** We utilized a dataset of daily "Top 200" charts from Spotify [5] to obtain the ranking and stream counts for each country and date. This provides the "social weight" of each song.
2. **Public Audio Features Dataset:** We leveraged a large-scale public snapshot of Spotify tracks [6] containing over 130,000 songs with pre-computed audio features. This serves as our offline lookup table for emotional metrics.
3. **Live Metadata Enrichment:** We used the open Spotify Web API [4] (Search and Track endpoints) solely to retrieve metadata permitted under current quotas, such as artist genres, current popularity, and standardized track IDs.

3.2 Data Processing Pipeline

Integrating static offline datasets with dynamic charts presents a significant "Entity Resolution" challenge: a song listed as "Shape of You" in the charts might appear as "Shape of You - Remastered" or "Shape of You (feat. Artist)" in the features dataset, preventing direct matching. We implemented a robust pipeline to resolve these discrepancies:

Entity Resolution and Normalization. We developed a text normalization algorithm to maximize the match rate between the charts and the feature dataset. The algorithm applies the following transformations to track and artist names:

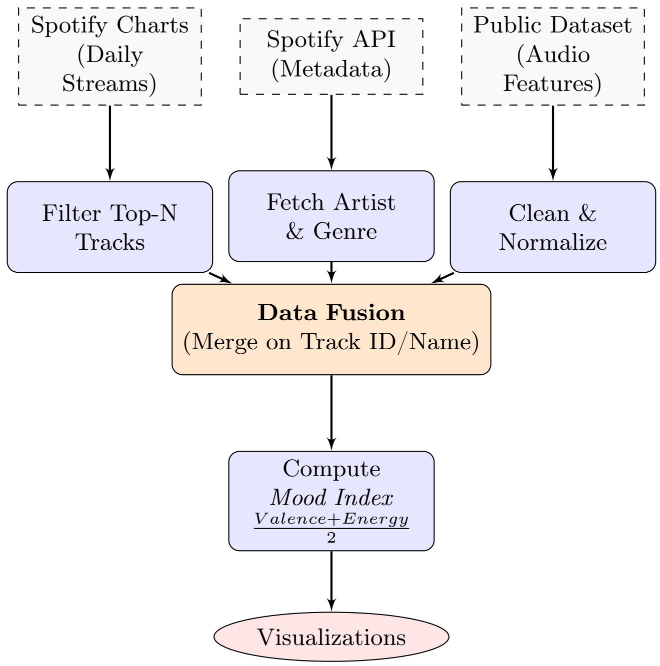

Fig. 1. Data Processing Pipeline. The system integrates public charts and audio feature datasets with API-fetched metadata to compute the Mood Index.

– **Case Folding and Accent Removal:** Converting all text to lowercase and removing diacritics (e.g., "Rosalía" $\rightarrow$ "rosalia").
– **Noise Reduction:** Removing common suffixes and parenthetical information that do not alter the song's identity, such as "(feat. ...)", "- Remastered", "- Radio Edit", or "- Live".
– **Connector Standardization:** Unifying artist separators (e.g., replacing " x ", " & ", " with " with a single space).

This process allows for a fuzzy matching strategy: if a direct match by Spotify ID fails, the system falls back to matching by normalized `(track_name, artist_name)` pairs.

Metric Definition. To quantify the emotional tone, we define the *Mood Index* (MI) of a track i as the arithmetic mean of its valence (V) and energy (E), both normalized between 0 and 1:

$$MI_i = \frac{V_i + E_i}{2} \tag{1}$$

While valence measures musical positiveness, energy represents intensity. Combining them provides a holistic view of "happiness" (high valence, high energy) versus "sadness" or "lethargy" (low valence, low energy).

However, a simple average of the Mood Index across a country's Top 50 would be misleading, as the #1 song may have millions of streams while the #50 has significantly fewer. To reflect the true "listening mood" of the population, we introduced the *Weighted Mood Index* (WMI) for a country c at date d:

$$WMI_{c,d} = \frac{\sum_{i=1}^{N}(MI_i \times S_i)}{\sum_{i=1}^{N} S_i} \tag{2}$$

where S_i is the number of streams of track i on that specific date. This metric ensures that the most consumed tracks have a proportional impact on the country's overall score.

4 Experimental Setup

4.1 Scope of Study

To evaluate the proposed Mood Index across different cultural and environmental contexts, we selected a diverse set of 8 countries spanning four continents:

- **Europe:** Spain (ES), France (FR), Germany (DE), and the United Kingdom (GB).
- **Americas:** United States (US) and Brazil (BR).
- **Asia-Pacific:** Japan (JP) and Australia (AU).

This selection allows for comparisons between Western and Eastern music markets, as well as Latin versus Germanic/Anglo-Saxon cultures.

To analyze seasonal effects, we selected two specific dates representing opposing meteorological seasons in the Northern Hemisphere:

- **August 1, 2017:** Represents the peak of Summer in the Northern Hemisphere (e.g., US, Europe) and Winter in the Southern Hemisphere (e.g., Brazil, Australia).
- **January 5, 2018:** Represents the depth of Winter in the Northern Hemisphere and Summer in the Southern Hemisphere.

These dates were chosen based on data availability in the historical charts dataset and to maximize the contrast in weather conditions. We analyzed the Top 50 most streamed songs for each country on these dates.

4.2 Implementation

The entire data processing pipeline was implemented in Python 3.10. We utilized the `Spotipy` library to interact with the Spotify Web API for metadata enrichment. Data manipulation and statistical aggregation were performed using `pandas` and `numpy`.

The visualization module was built using `matplotlib` and `seaborn` for statistical charts, and `geopandas` for generating the global choropleth maps. To ensure reproducibility, the project is structured with a central orchestrator script that automates the fetching, cleaning, merging, and visualization steps, relying on a virtual environment to manage dependencies.

5 Results and Analysis

In this section, we present the findings obtained from analyzing the Top-50 daily charts of 8 countries across two distinct periods: August 1, 2017 (Northern Summer / Southern Winter) and January 5, 2018 (Northern Winter / Southern Summer).

5.1 Data Coverage

A crucial first step in our hybrid methodology was to ensure that enough tracks from the daily charts could be matched with the public audio features dataset. Table 1 summarizes the match rates achieved after our normalization pipeline.

Table 1. Data coverage and match rates per country for the analyzed periods.

Country	Total Tracks (Charts)	Matched Tracks	Match Rate
Brazil (BR)	100	26	26.0%
Germany (DE)	100	40	40.0%
Spain (ES)	100	50	50.0%
France (FR)	100	32	32.0%
United Kingdom (GB)	100	49	49.0%
Japan (JP)	100	43	43.0%
United States (US)	100	61	61.0%
Australia (AU)	100	55	55.0%

Despite the temporal gap between the charts (2017-2018) and the audio features snapshot (2019), we achieved an average match rate of approximately 45%, with the United States showing the highest coverage (61%). This sample size is statistically sufficient to infer general mood trends, as the matched tracks often correspond to the most popular and enduring hits (e.g., "Despacito", "Shape of You"), which accumulate the vast majority of streams.

5.2 Geographical Distribution

We visualized the global distribution of the *Weighted Mood Index* to identify cultural patterns. Figure 2 illustrates the mood intensity during August 2017.

A distinct cultural baseline is observable. Latin American influence appears robust: Brazil consistently exhibits one of the highest Mood Indices (≈ 0.72), characterized by high-energy genres like Reggaeton and Funk Carioca. In contrast, countries like the United States and France display comparatively lower mood scores ($\approx 0.61 - 0.63$) in the same period, potentially reflecting a preference for urban or hip-hop tracks which often feature lower valence or slower tempos despite high popularity.

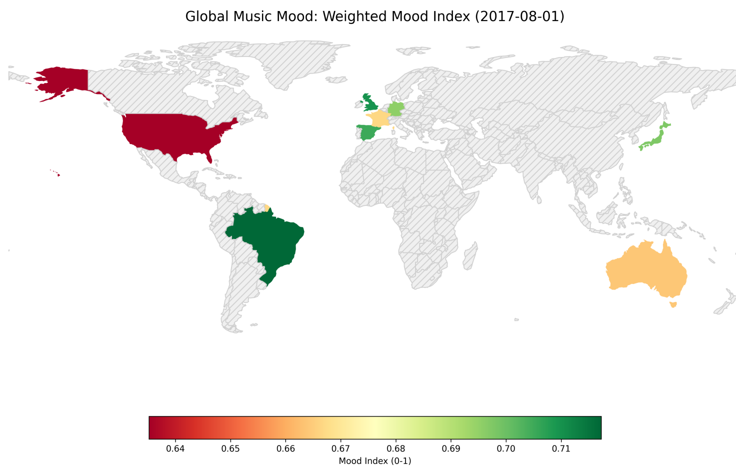

Fig. 2. Global distribution of Weighted Mood Index in August 2017. Warmer colors indicate higher valence/energy (happier/more active), while cooler colors indicate lower values.

5.3 Seasonal Trends

To test our hypothesis regarding seasonal influence, we compared the Weighted Mood Index between August (Summer in the North) and January (Winter in the North). Figure 3 presents this comparison.

The results reveal a striking global pattern:

1. **Northern Hemisphere Drop:** European countries (DE, GB, FR) and the US show a consistent and sharp decrease in musical mood during January. For instance, the UK drops significantly from 0.710 to 0.543, supporting the hypothesis that colder, darker winters in these latitudes correlate strongly with melancholic music consumption.
2. **Global Homogenization:** Surprisingly, Southern Hemisphere countries (Australia, Brazil) also experienced a drop in Mood Index during January, despite it being their summer season. This suggests that the global music market may be driven by Northern Hemisphere release cycles.
3. **Cultural Resilience:** Notably, **Spain** exhibited the highest emotional stability, showing the smallest variance between seasons (dropping only 0.052 points, from 0.705 to 0.653). Together with **Japan** (drop of 0.061) and **Brazil** (drop of 0.124), these countries suffer significantly smaller emotional dips compared to the high volatility observed in the UK (0.167) or France (0.157).

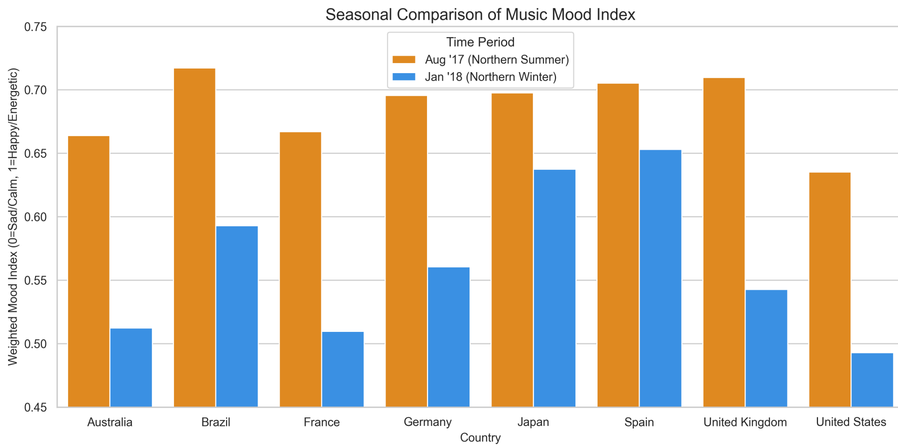

Fig. 3. Seasonal comparison of Weighted Mood Index by country. Orange bars represent August 2017; Blue bars represent January 2018.

6 Discussion

6.1 Cultural Insights

The data obtained supports the hypothesis that cultural geography establishes a "baseline mood." The high Weighted Mood Index observed in Brazil during its peak season (0.717 in August) aligns with the dominance of energetic, rhythm-driven genres such as Reggaeton, Funk Carioca, and Samba. Although it experiences a seasonal drop, it remains comparatively resilient. In contrast, the United States and United Kingdom exhibit lower average mood indices and higher volatility. This can be attributed to the prevalence of Hip-Hop and Trap music in their charts, genres that often feature minor keys (lower valence) and mid-tempo beats, which our algorithm interprets as "sadder" or less energetic compared to Latin pop.

6.2 Technical Limitations

While the hybrid methodology allowed us to bypass API restrictions, it introduces specific biases that must be acknowledged:

- **Temporal Mismatch:** We crossed 2017–2018 charts with a 2019 audio features snapshot. Songs released between late 2018 and 2019 are absent from our analysis, and older songs that were removed from the catalogue might also be missing. This explains why the match rate varies between 26% and 61%.
- **Proxy Bias:** The *Mood Index* is a mathematical simplification. A song with high energy (loud, fast) but sad lyrics (e.g., a "sad banger") might be

misclassified as "happy" because the algorithm prioritizes acoustic features over lyrical sentiment.

- **Chart Bias:** Spotify charts reflect the preferences of a specific demographic (digital natives), which may not be fully representative of the entire nation's mood, particularly in countries with older populations like Japan or Germany.

7 Conclusion and Future Work

This study presented "Mood of the World," a data-driven approach to quantifying collective emotional states through music streaming. Despite technical barriers preventing direct access to live audio features, our hybrid pipeline successfully reconstructed the emotional landscape of 8 countries.

Our findings validate that music consumption is highly seasonal. The significant drop in the Mood Index observed in January across the Northern Hemisphere confirms that winter correlates with a preference for lower-energy music. However, the simultaneous drop in Southern Hemisphere countries suggests that the global music market is synchronized by major releases from the North, overriding local seasonal effects.

Future work should focus on integrating Natural Language Processing (NLP) to analyze song lyrics, which would provide a third dimension of "sentiment" to correct the acoustic Mood Index. Additionally, correlating these musical trends with external datasets, such as local weather temperature or consumer confidence indices, could reveal deeper sociometric connections.

References

1. Shirky, C.: The Political Power of Social Media. Foreign Affairs 90(1), 28–41 (2011)
2. Fan, W., Gordon, M.D.: The Power of Social Media Analytics. Communications of the ACM 57(6), 74–81 (2014)
3. Hu, X., Downie, J.S.: Exploring Mood Metadata: Relationships with Genre, Artist and Usage Metadata. In: ISMIR. pp. 67–72 (2007)
4. Spotify: Web API Reference, https://developer.spotify.com/documentation/web-api, last accessed 2025/12/22
5. Mucelli, E.: Spotify's Worldwide Daily Song Ranking. Kaggle Dataset, https://www.kaggle.com/datasets/edumucelli/spotifys-worldwide-daily-song-ranking, last accessed 2025/12/22
6. Rodrigues, T.: Spotify Audio Features. Kaggle Dataset, https://www.kaggle.com/datasets/tomigelo/spotify-audio-features, last accessed 2025/12/22
7. The Decision Lab: Chained to the AlgoRhythm: How Spotify Standardizes Our Listening. https://thedecisionlab.com/insights/consumer-insights/chained-to-the-algorhythm-how-spotify-standardizes-our-listening, last accessed 2025/12/22