

Data Set Properties

NAME:

"Billboard Music Genre Over Time"

FROM:

<https://github.com/kevinschaich/billboard/tree/master>

(<https://github.com/kevinschaich/billboard/tree/master>)

SIZE:

The dataset size is dynamic and is built over entries of music tracks over time.

TIMEFRAME:

It is a constantly evolving dataset capturing real-time data as it spans across various years with entries such as from 2017.

UNIQUE COLUMNS:

7 (date, year, broad_genre, loudness, duration_ms, lyrics, day, month)

General Summary:

The dataset encompasses entries detailing various aspects of music tracks that made it to the Billboard charts. Each row captures the unique identity of the track including its date of entry into the chart, genre classification, attributes of the audio like loudness and duration in milliseconds, and the lyrics of the song. Additionally, the dataset is enriched with separated day and month columns for more granular time analysis.

Specific Column Descriptions:

- **Date** (*Interval*): The date when the song was entered or ranked on the Billboard charts.
- **Year** (*Interval*): The year of the song's entry onto the

charts.

- **Broad_Genre** (*Categorical*): The broad genre classification of the song like pop, rap, r&b, rock, etc.
- **Loudness** (*Ratio*): The loudness of the track measured in decibels (dB).
- **Duration_ms** (*Ratio*): The total duration of the song measured in milliseconds.
- **Lyrics** (*Text*): The lyrics of the song, a rich text data for linguistic and sentiment analysis.
- **Day** (*Ordinal*): The day of the month when the song made it to the Billboard charts.
- **Month** (*Ordinal*): The month of the year when the song entered the Billboard charts.

Goals in Place for this Data:

- **Trend Analysis:** Analyzing the trends in music genres over the years, providing insights into the evolution of musical taste and popularity of genres.
- **Audio Attributes Study:** Exploring how audio attributes like loudness and duration correlate with the popularity or ranking of the songs.
- **Lyric Analysis:** Utilizing the lyrics data for sentiment analysis, linguistic patterns, and correlating it with genres or song popularity.
- **Time Series Analysis:** Employing the day, month, and year data to study the temporal patterns in song entries, genres, and other attributes on the Billboard charts.

Sample Data Descriptions:

1. Date:

- *char*: The date of the song's entry onto the Billboard chart, formatted as MM/DD/YY.

2. Year:

- *int*: The specific year of the song's placement on the

Billboard chart.

3. **Broad_Genre:**

- *char*: Indicates the broader genre category to which the song belongs, e.g., pop, rap, rock.

4. **Loudness:**

- *float*: A numerical value indicating the loudness of the song, measured in decibels.

5. **Duration_ms:**

- *int*: Indicates the total duration of the song in milliseconds.

6. **Lyrics:**

- *char*: Contains the full lyrics of the song, offering opportunities for textual and sentiment analysis.

7. **Day and Month:**

- *int*: These fields provide a more detailed temporal context for each song's placement on the chart, aiding in more granular time series analysis.

Example Uses:

This dataset is a goldmine for music analysts, linguistic researchers, and cultural study enthusiasts. It can be used to correlate musical attributes with popularity, study the evolution of genres over time, and conduct in-depth lyrical analyses to extract themes, sentiments, and linguistic patterns prevalent in popular music across different time periods.

Source:

The dataset is processed internally, with entries spanning multiple years, offering a rich source for historical and trend analysis in the realm of popular music. The data's granularity and richness make it a valuable resource for a wide array of analytical explorations.