

Spotify Track Analysis

Yoatam Gebremicael

Problem

Details on the Problem:

This analysis focuses on how different audio features related to songs like energy, danceability, and mood relate to a song's popularity. Further, I will check whether a highly popular song present on an album tends to influence the popularity of other tracks present on the same album. I will also check if collaborations between artists result in unique characteristics, such as more energy or increased danceability, compared to solo tracks. This may help determine a pattern of the preferences of listeners and track performances that could help Spotify improve engagement.

Research Questions:

1. How does the presence of one or more highly popular songs in an album affect the popularity and listener engagement of other tracks within the same album?

- **Sub-question:** Do high-popularity albums show higher listener retention for less popular tracks?
- **Sub-question:** Are characteristics that make tracks popular (high energy or danceability) driving more interaction with other tracks on that album that share those qualities?

2. How do the musical attributes of tracks change when artists collaborate compared to when they work solo?

- **Sub-question:** Do the collaborations significantly show consistent change within certain musical qualities, such as tempo, loudness, and energy, across genres? For instance, are pop collaborations generally more danceable whereas rock collaborations are louder and have higher intensities?
- **Sub-question:** Does the popularity of collaboration differ across genres, perhaps because collaboration tracks are preferred over solo performances in genres?

Impact

Institutional Impact:

By understanding the impact of popular songs within an album and the effects of collaborations on song popularity, Spotify can enhance its recommendation system, and playlist curation, and take advantage of popular songs to drive listener engagement even higher.

Benefits:

Improved Recommendations: Identifying which audio features are most important to popularity could allow Spotify to improve its recommendation algorithms, suggesting tracks with user-preferred qualities, such as high-energy or highly danceable songs.

Customized Collaborative Playlists: Knowing how musical attributes shift in collaborations (such as increased danceability or energy) can be used by Spotify in generating playlists that match a particular attribute. This could be the case if collaborations had higher energy; then, Spotify can make playlists for those who want higher levels of dynamism and energy when listening, thereby personalizing playlist creation to listener preference.

Data

Dataset Size:

The dataset consists of 114,000 tracks across 114 genres.

Data Collection:

The data was collected and cleaned using Spotify's Web API, ensuring the correctness of the track and its audio features.

Cleanliness of the Data:

Most columns are almost complete in this dataset, with minor missed values in artist, album_name, and track_name columns. In preparing this data for analysis, such missing values should be dealt with. Besides, it will be important to convert the

duration_ms from millisecond values to minutes for better readability and comparison in analysis.

Additional Variables for Consideration:

- **User Listening History:** Adding metrics like track replays could help explain why specific songs draw users back more than can be explained by pure popularity.

Methods

Key Variables

- **Popularity:** Refers to the performance-based measure of a track, often by the number and recency of plays.
- **Audio Features:** The various attributes that include danceability, energy, valence, tempo, and loudness make up vital qualities such as intensity, positivity, speed, and volume in understanding track popularity.
- **Collaborations:** Comparing tracks with multiple artists against solo performances may show how collaborations influence audio features and popularity.

New Variables:

Creating a popularity index ("low," "medium," and "high") can help determine how trends in audio features change across these three levels.

Visualizations and Summary Statistics:

- **Popularity Impact:** A box plot that compares the popularity of tracks in albums with one or more highly popular songs to those without.
- **Collaborative vs. Solo tracks:** A grouped bar chart that compares average danceability and energy levels in collaborative versus solo tracks.

- **Energy vs. Popularity:** A Histogram that shows the distribution of energy and danceability levels across both collaborative and solo tracks.

Key Visualization:

- **Popularity Impact:** A box plot comparing album track popularity for albums that have one or more very popular songs versus albums that do not will help to assess how the presence of popular tracks affects the overall engagement and popularity of other songs in the same album.
- **Collaborative vs Solo Tracks:** A grouped bar chart showing average danceability and energy levels for collaborative tracks and solo works will give insight into how collaborations typically affect these attributes.

Concerns

Data Integrity:

Track popularity can vary according to various reasons, including the presence of a particular social media trend or promotion by an artist that is not included in this dataset. Because of this, misleading conclusions may arise also due to the negligence of such dynamics.

Possible Limitations of the Analysis:

- **Popularity vs. Preference:** In Spotify, popularity is based on play counts, which does not signify the listener's satisfaction or indicate long-term listening.
- **External Influences:** Social media, promotions by artists, and other exogenous factors that can drastically change track popularity are not represented in the dataset. Without these variables, any conclusions about popular songs driving other album tracks or characteristics unique to collaboration may also be biased.