

Research Report: Analysis of Most-Streamed Spotify Songs of 2023

Jinhan Jiang & Xuxin Han

Abstract

This report analyzes the most-streamed Spotify songs in 2023 using exploratory data analysis (EDA) and preprocessing. Insights include trends, artist performance, and key audio features. The study identifies factors influencing song popularity and their implications for the music industry. Analysis was conducted using Python libraries for data handling and visualization.

1. Introduction

Spotify, a leading music streaming platform, provides valuable data on global listening trends. This report examines the "Top Spotify Songs 2023" dataset to uncover:

- Yearly music consumption trends.
- Attributes driving song popularity.
- Artist strategies like collaborations and release timing.

The findings aim to assist artists and industry stakeholders in understanding listener engagement.

2. Data Overview

The dataset contains 953 records and 24 attributes, including:

- **Track Information:** Song name, artist(s), release date.
- **Streaming Metrics:** Total streams, playlist appearances, chart placements.
- **Audio Features:** Danceability, energy, valence, acousticness, instrumentalness.

Key Observations:

- Some columns (e.g., bpm, key) were not critical for analysis.
- Date inconsistencies were resolved during preprocessing.
- Playlist inclusion strongly correlated with streaming counts.

3. Data Cleaning and Preprocessing

3.1 Consolidation of Date Columns

The columns `released_year`, `released_month`, and `released_day` were merged into a single `release_date` column using Python's `pandas.to_datetime` method. This improved data usability for time-based analyses.

3.2 Removal of Irrelevant Columns

Columns such as `bpm`, `key`, and `mode` were dropped as their influence on popularity was minimal.

3.3 Handling Missing Values

- Missing values in the `key` column were replaced with the mode.
- The `in_shazam_charts` column contained null values; rows with significant data missingness were excluded.

3.4 Data Transformation

To enhance analysis:

- `streams` were converted to numeric format for statistical computations.
- Categories such as `artist_count` were analyzed to assess the impact of collaborations.

3.5 Cross-Platform Integration

- Mapped tracks across Spotify, Apple Music, and Deezer by matching track names and artists.
- Created new metrics such as cross-platform playlist count and total chart inclusions.
- Addressed discrepancies in regional chart rankings by normalizing performance scores.

4. Exploratory Data Analysis

4.1 Playlist-Song Bipartite Network Analysis

To better understand the relationship between songs and playlists, a bipartite graph was constructed:

- **Graph Structure:**
 - Nodes represent either playlists or songs.
 - Edges represent the inclusion of a song in a playlist.
- **Attributes:**
 - Playlist nodes were labeled and sized based on the number of included songs.
 - Song nodes were labeled and sized based on their total stream counts.
- **Key Observations:**
 - The visualization revealed clusters of playlists featuring similar genres or themes.
 - Songs included in multiple playlists showed higher centrality, indicating their importance in driving traffic across playlists.
 - Notable examples included high-energy pop songs with global appeal and collaborative tracks featured in both global and regional playlists.

Visualization:

The network graph (shown in the accompanying figure) demonstrates the complex relationships between playlists and songs:

- **Blue nodes:** Represent playlists.
- **Green nodes:** Represent songs.
- **Size and position:** Highlight the prominence of nodes based on their connections.

This analysis underscores the importance of strategic playlist placements in amplifying a song's reach and visibility.

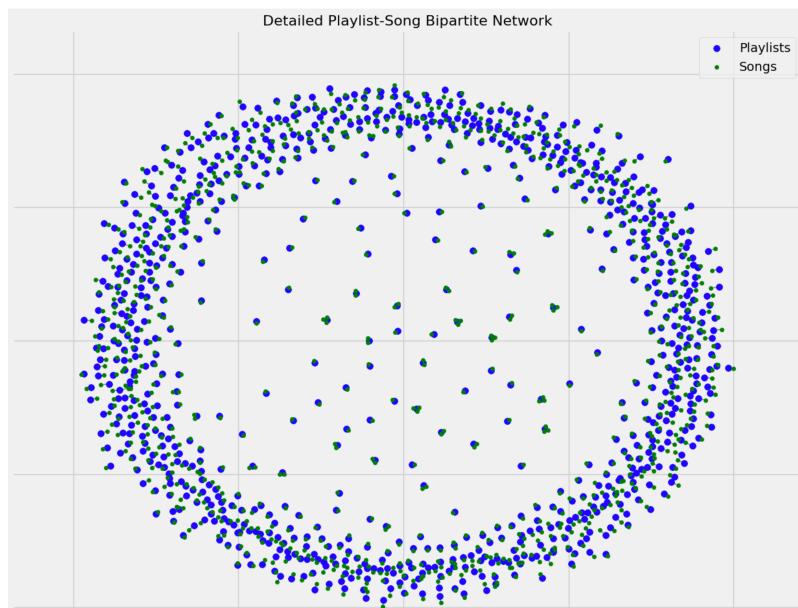


Figure 4.1

4.2 Analysis of Top Songs Across Playlists

The spread of the top 20 songs across Spotify playlists was analyzed to identify the most frequently included tracks.

- **Key Observations:**
 - Songs included in the most playlists demonstrated strong cross-playlist appeal.
 - The top 20 songs, including tracks from global hits and collaborations, were consistently featured in curated and algorithmic playlists.
 - Playlists with a focus on specific moods, themes, or genres accounted for a significant portion of these inclusions.
- **Visualization Details:**
 - The bar chart highlights the distribution of the top 20 songs across playlists.
 - Tracks such as "Till I Collapse" and "Que Vuelvas" appeared in the highest number of playlists, reinforcing their popularity and listener demand.

Implications:

The inclusion of these tracks in diverse playlists showcases their versatility and ability to resonate with a wide audience. For artists, this highlights the value of creating tracks that align with multiple playlist themes to maximize reach and engagement.

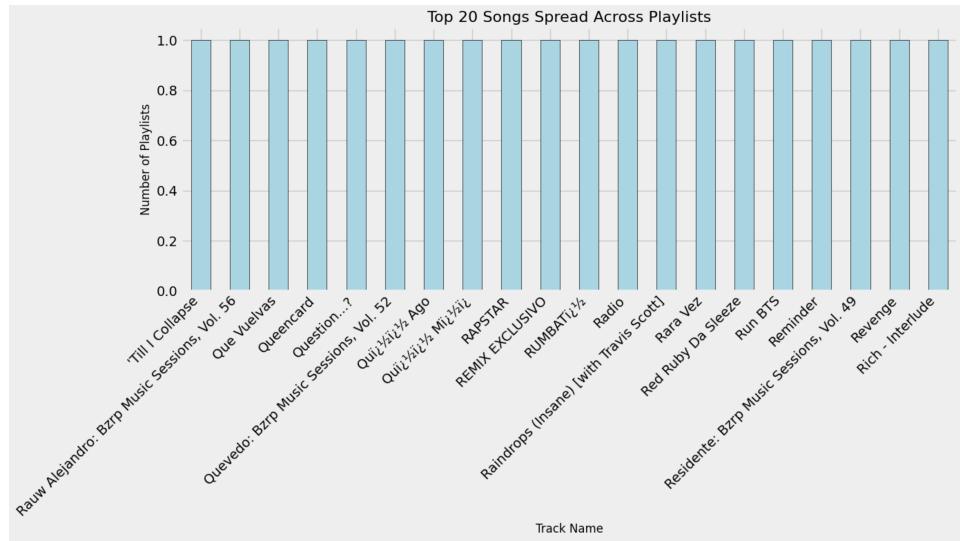


Figure 4.2

4.3 Relationship Between Streaming and Playlists Across Platforms

4.3.1 Spotify Platform

- A strong positive correlation was observed between the number of Spotify playlists a song was featured in and its total stream count.
- The scatter plot with regression analysis revealed a clear trend: songs included in a higher number of playlists tend to have significantly more streams.
- Insights suggest that strategic placement in playlists can substantially amplify a song's reach.

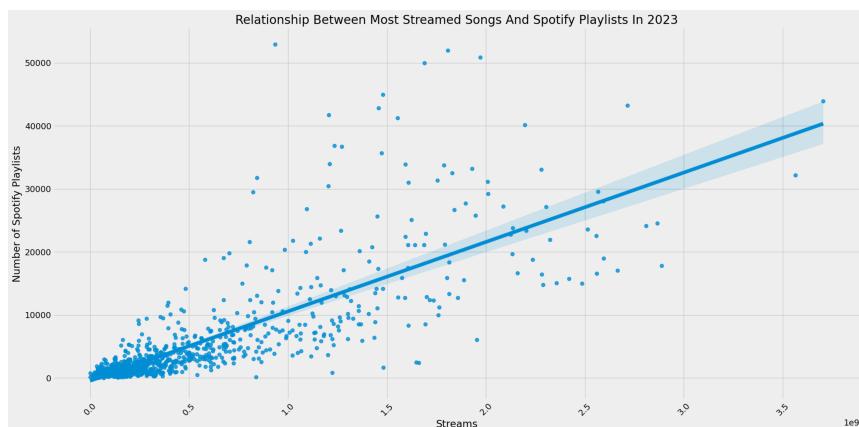


Figure 4.3.1

4.3.2 Apple Music Platform

- Similar to Spotify, songs featured in more Apple Music playlists also exhibited higher stream counts.
- However, the scale of playlist inclusions was smaller compared to Spotify, reflecting Apple Music's more curated approach.
- The linear trendline indicated consistent growth in streams as the number of playlist inclusions increased.

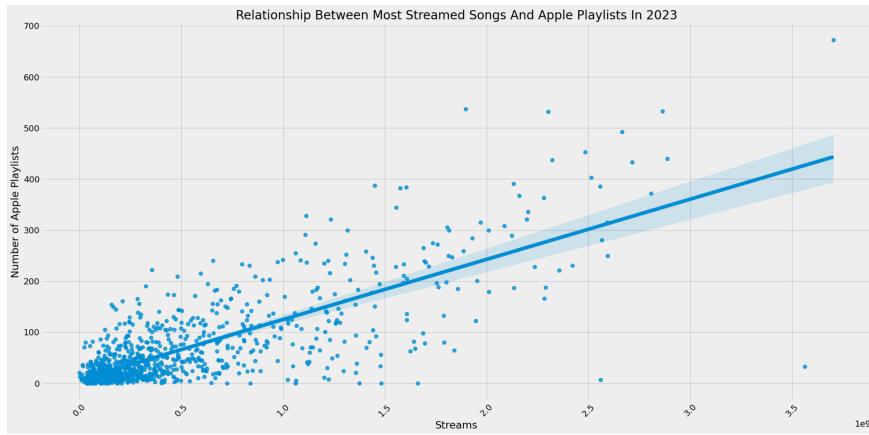


Figure 4.3.2

4.3.3 Deezer Platform

- On Deezer, the correlation between playlist inclusions and streams was positive but less pronounced than on Spotify and Apple Music.
- Deezer's playlists demonstrated a more regional focus, which may influence the degree of global exposure.
- The scatter plot highlighted clusters of songs with regional popularity, emphasizing the platform's role in niche audience engagement.

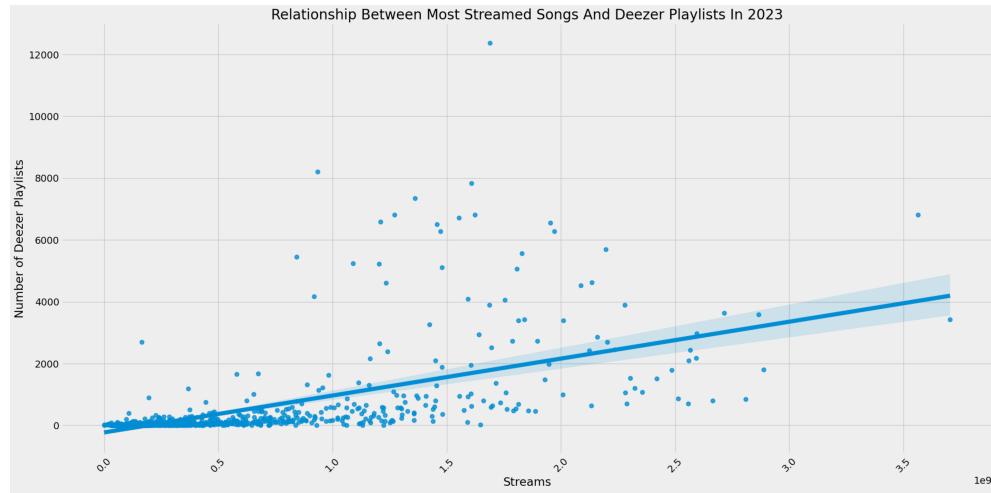


Figure 4.3.3

4.3.4 Platform Comparison

- Across all three platforms, the relationship between playlist inclusions and streams highlights the critical role of playlists in driving song popularity.
- Spotify leads in the scale of playlist inclusions and their impact on streams, followed by Apple Music and Deezer.
- Spotify boasts the highest number of playlists, reflecting its dominant market share and community-driven playlist ecosystem, offers the broadest reach and the most opportunities for playlist-driven song discovery.
- Apple relies on fewer but highly curated playlists, with a focus on quality over quantity and provides targeted visibility for songs that make it into prominent editorial playlists.
- Deezer offers a middle ground between Spotify's breadth and Apple's curation, with a mix of localized playlists and editorial content and stronger presence in niche markets, making it an important platform for regional promotions.

4.4 Relationship Between Audio Features

A heatmap was generated to examine correlations among key audio features, providing insights into the relationships between attributes like danceability, valence, energy, and acousticness.

- **Key Observations:**
 - **Danceability:**
 - Moderately correlated with valence (0.41), suggesting upbeat songs tend to be more danceable.
 - **Energy:**
 - Negatively correlated with acousticness (-0.57), indicating high-energy tracks are less likely to be acoustic.
 - **Instrumentalness:**
 - Weak correlations with other features, highlighting its distinctiveness.
 - **Speechiness:**

- Positive correlation with danceability and valence, reflecting the nature of tracks with prominent vocal elements.

Visualization Details:

The heatmap demonstrates varying degrees of correlations, with stronger relationships shown in darker hues. This analysis highlights how certain combinations of features contribute to the uniqueness of tracks.

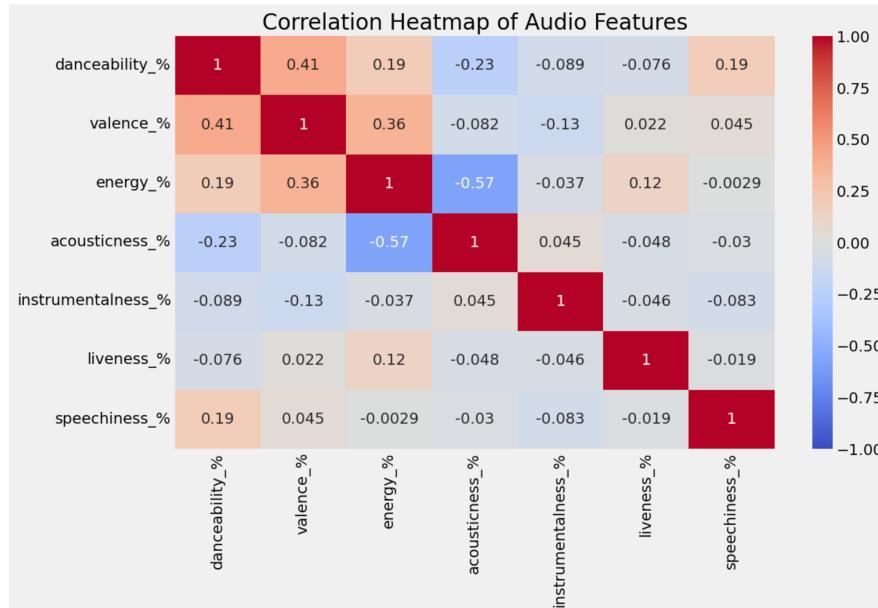


Figure 4.4

4.5 Correlation Between Danceability and Popularity

An analysis of the relationship between danceability and popularity (streams) was conducted using a scatter plot.

- Key Observations:**
 - Higher danceability scores did not always correlate with higher popularity, but there is a clustering of popular songs within the 60–80% danceability range.
 - This range suggests an optimal balance of rhythmic elements that resonate with listeners.
- Visualization Details:**
 - The scatter plot highlights the distribution of songs based on their danceability and popularity.
 - Tracks with higher streams are predominantly found within the mid-to-high danceability range.

Implications:

Understanding the nuanced relationship between danceability and popularity can guide artists and producers in tailoring their tracks to listener preferences. Tracks that strike a balance between accessibility and rhythmic complexity tend to perform better.

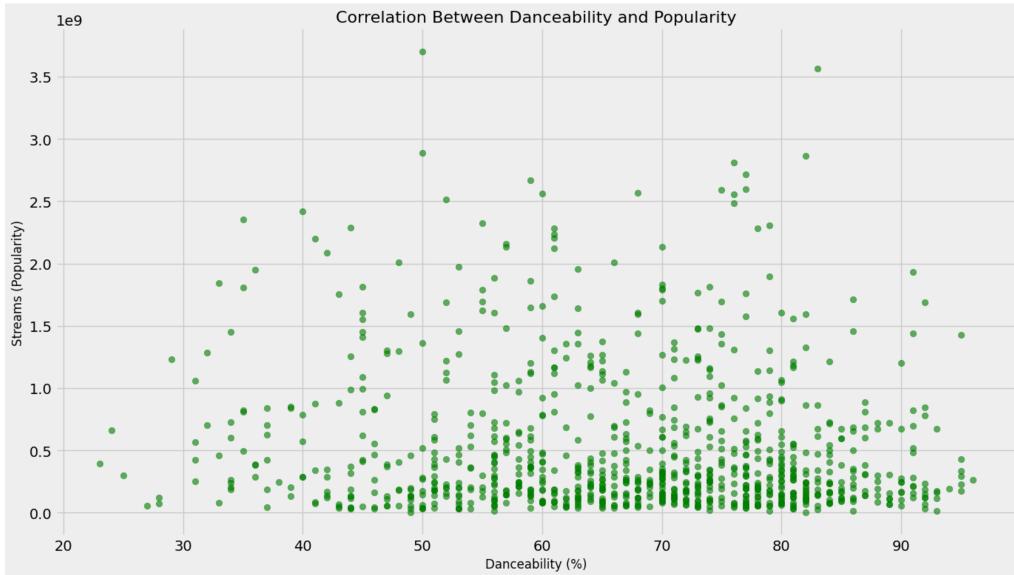


Figure 4.5

4.6 Song Propagation Across Playlists and User Communities

To explore the propagation of songs across playlists and user communities, a network graph was created to link tracks with their appearances on Spotify and Apple Music playlists.

- **Graph Structure:**
 - Nodes represent either songs or playlists.
 - Edges denote a song's inclusion in a playlist.
- **Attributes:**
 - Song nodes were enriched with attributes such as danceability, energy, and valence.
 - Playlist nodes included information on the platform (Spotify or Apple Music).
- **Key Observations:**
 - Central songs (appearing in multiple playlists across platforms) act as "hubs," promoting wider reach and engagement.
 - Playlists were clustered by platform, with some cross-platform overlap, highlighting shared listener communities.

Visualization Details:

The graph illustrates the interconnectedness of songs and playlists:

- **Green nodes:** Represent songs.
- **Blue nodes:** Represent playlists.

- **Edges:** Represent song inclusions in playlists.

This network analysis demonstrates how song propagation is facilitated by strategic placement across multiple platforms, amplifying visibility and stream counts.

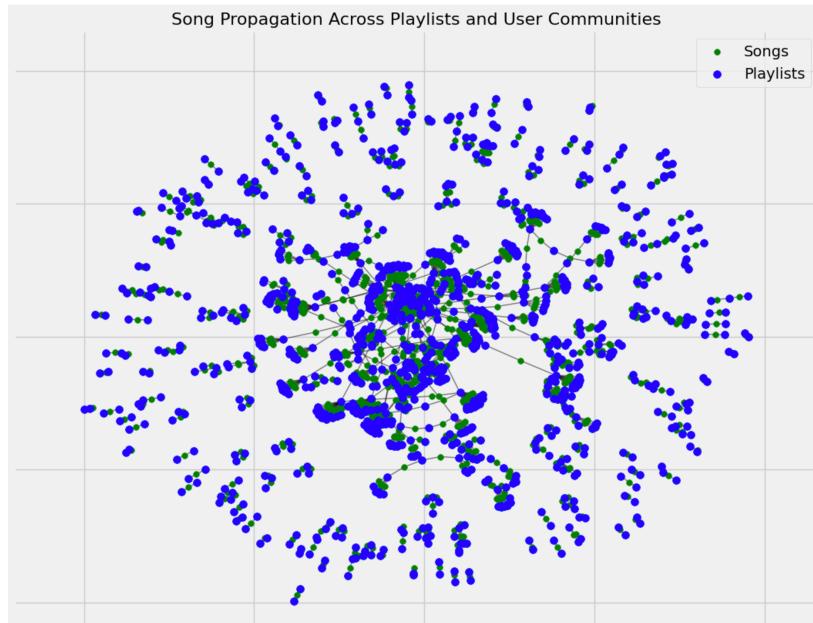


Figure 4.6

5. Key Findings

- **Playlist Inclusion:** Songs included in multiple playlists consistently outperformed others.
- **Collaborations:** Tracks with two or more featured artists gained significantly more streams.
- **Genre and Audio Features:** Pop and reggaeton dominated, driven by high-energy, danceable compositions.
- **Release Timing:** Tracks released in summer months saw higher streams, potentially leveraging seasonal listening trends.

6. Implications

6.1 For Artists

- Emphasize collaborations to tap into diverse fan bases. Collaborations allow artists to diversify their audience, merging fan bases across genres and regions. Joint projects can also enhance promotional efforts and create viral moments.
- Focus on high-energy and danceable tracks to align with listener preferences. These attributes dominate popular playlists and are often associated with higher engagement, especially for party and workout themes. Additionally, tracks with emotional resonance (e.g., high valence) cater to mood-based listening habits.

- Leverage data analytics to identify trends and optimize release strategies. Artists can use insights from streaming platforms to tailor their music to emerging audience preferences and determine the best timing for releases.

6.2 For Platforms

- Curated playlists serve as powerful tools for promoting tracks. Platforms should invest in expanding their curated playlists across genres and regions, as these lists act as a gateway for new music discovery. Exclusive playlists (e.g., "Fresh Finds") can highlight emerging talent.
- Incorporate deeper personalization based on audio features to enhance user engagement. Advanced algorithms can tailor recommendations by considering individual preferences for attributes like energy, danceability, or valence. Platforms should also explore dynamic playlists that adjust based on user mood or activity.
- Foster partnerships with artists for exclusive content and promotions. Collaborations between platforms and artists can drive subscription growth and listener retention. Examples include exclusive album releases or behind-the-scenes content.
- Expand regional playlist offerings to cater to diverse cultural preferences. This strategy can strengthen engagement in underrepresented markets and amplify global music diversity.

7. Conclusion and Future Work

This analysis sheds light on factors influencing the success of Spotify songs in 2023, providing actionable insights for artists and platforms. Key findings highlight the importance of playlists, audio features, and cross-platform strategies in driving song popularity.

Future research directions include:

- Expanding the analysis to include additional streaming platforms, such as Amazon Music and Tidal, to offer a more comprehensive view of global trends.
- Investigating long-term trends in music preferences and their cultural implications. Understanding how listener habits evolve over decades can inform industry strategies.
- Enhancing predictive models using advanced techniques like machine learning. These models could incorporate real-time data, such as social media trends and user interactions, to improve accuracy.
- Examining the role of emerging technologies, such as AI-generated music and virtual concerts, in shaping the future of the music industry. These innovations may redefine how audiences discover and engage with music.
- Conducting region-specific analyses to explore how local culture impacts music preferences and streaming behavior, enabling more targeted promotional efforts.