"Melody Metrics: Unveiling Patterns in Music Consumption"

Muhammad Sumeer Ahmad[†], Muhammad Farrukh[†]

[†]Stony Brook University

I. INTRODUCTION

The landscape of music consumption has been dramatically reshaped by streaming services, which have become primary platforms for music distribution and discovery. Spotify, Apple Music, Deezer, and Shazam lead the pack in transforming how songs are shared, listened to, and valued by audiences worldwide. With this shift, an abundance of data has become available, presenting new opportunities to analyze musical trends and industry dynamics. This project proposes to delve into a comprehensive dataset from Spotify that captures the essence of music consumption in 2023, using a variety of attributes and metrics to paint a detailed picture of the current music scene.

The richness of the Spotify dataset provides a unique lens through which to view the interactions between music tracks, artists, and listeners. By examining this dataset, we can uncover patterns in musical preferences, the impact of artists on track popularity, and how songs traverse various digital platforms. This kind of analysis not only enhances our understanding of digital music trends but also assists stakeholders in the music industry—ranging from artists to executives—in making informed decisions that align with listener preferences and emerging trends.

II. DATASET SPECIFICATION

The dataset [1] under consideration offers an extensive collection of metrics and features from the most popular songs on Spotify in 2023. It includes detailed information on each track, such as the track name, artist names, release dates, and the song's presence across several key music platforms. Additionally, the dataset is enriched with data points on streaming statistics, playlist inclusions, chart positions, and comprehensive audio features like tempo, key, and danceability. Each entry in the dataset is multi-dimensional, providing both quantitative and categorical data that reflect a song's commercial success and its musical characteristics.

Beyond basic track and artist information, the dataset also quantifies each song's integration into digital music ecosystems, with metrics such as the number of playlists a song appears in across Spotify, Apple Music, and Deezer, as well as its ranking on these platforms and on Shazam charts. The dataset's depth is further demonstrated by its inclusion of various percentages that describe the musicality of the tracks, such as danceability, valence, energy, acousticness, instrumentalness, liveness, and speechiness. This comprehensive dataset serves as the backbone for robust data-driven analysis in the project, aiming to explore multiple facets of music consumption and distribution in the digital age. Some important key features are:

- **streams:** Total number of streams on spotify in 2023
- **bpm:** Beats per minute, a measure of song tempo
- **mode:** Mode of the song (major or minor)
- **key:** key of the song (A-G)
- **danceability_%:** Percentage indicating how suitable the song is for dancing
- **valence_%:** Positivity of the song's musical content
- **energy_%:** Perceived energy level of the song
- **acousticness_%:** Amount of acoustic sound in the song
- **instrumentalness_%:** Amount of instrumental content in the song
- **liveness_%:** Presence of live performance elements
- **speechiness_%:** Amount of spoken words in the song
- **released_data:** Date when the song was released
- **artist_count:** Number of artists contributing to the song
- **in_spotify/apple/deezer/shazam_charts:** Presence and rank of the song on Spotify/Apple/Deezer/Shazam charts
- **in_spotify/apple/deezer_playlists:** Number of Spotify/Apple/Deezer Music playlists the song is included in

As our project evolves, we plan to enrich our dataset and deepen our analysis by integrating data directly from the Spotify API [2]. This powerful tool will allow us to supplement our existing data with a wealth of additional information such as genre, artist influences, related artists, and detailed user engagement metrics like saves and follows. For instance, we can use the Spotify API to dynamically categorize songs by genre based on their track names, enabling us to analyze trends and preferences within specific musical genres. Furthermore, the API's capability to fetch real-time data about artist popularity and listener habits will allow us to create interactive visualizations that show how artist popularity fluctuates over time or in response to certain events. This integration will not only enhance the robustness of our dataset but also enable more sophisticated, multi-dimensional visual analyses of music streaming patterns, offering insights that are both richer and more nuanced.

III. PROBLEM STATEMENT

This project is centered around the challenge of understanding the multifaceted nature of music popularity and its determinants in the era of streaming services. Despite the wealth of data available, deciphering the underlying patterns

that contribute to a song's success across various platforms remains complex. The problem is how to effectively analyze and interpret this vast array of data to yield actionable insights. Questions of particular interest include determining the factors that most significantly impact a song's popularity, how these factors vary across different platforms, and the evolution of musical preferences over time.

Furthermore, the project seeks to address how different audio features influence listener preferences and how these preferences translate into measurable outcomes like streaming numbers and chart positions. By tackling these questions, the project aims to provide a comprehensive analysis of current trends in the music industry, offering valuable insights into how artists and their music resonate with audiences in a digital context. The resulting analysis could serve as a foundational tool for industry stakeholders to forecast trends, tailor marketing strategies, and ultimately enhance the listener's experience. This problem statement sets the stage for a thorough exploration of the dataset and contributes to the broader understanding of digital music consumption dynamics.

IV. APPROACH

Keeping in mind our problem statement, we propose possible visualizations to achieve our objectives:

A. Music Analysis

To explore patterns in audio features to understand trends and preferences in popular songs, A **Scatter Plot Matrix** can be used. It will show relationships between different audio features, such as danceability, energy, and valence.
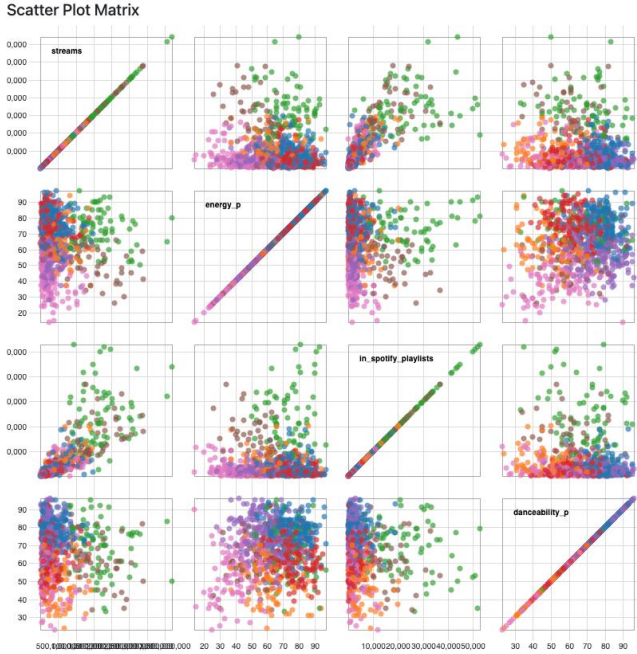

Fig. 1. scatter plot matrix

B. Platform Comparison

To Compare the song's popularity across different music platforms. We can utilize the following visualizations:

- **Pie Charts:** Show popularity (e.g., streams, downloads) of songs across different platforms Fig[2].
- **Line Graphs:** Compare changes in popularity over time across platforms Fig[3].
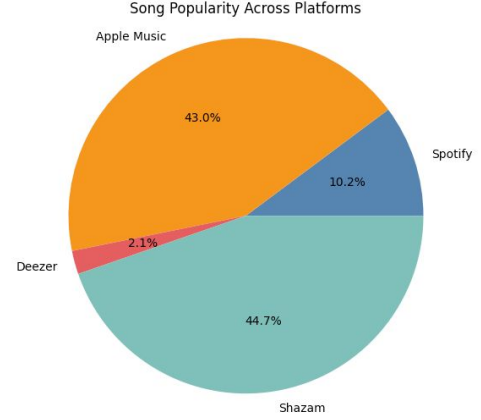- **Stacked Area Charts:** Illustrate the proportion of total streams per platform over time.
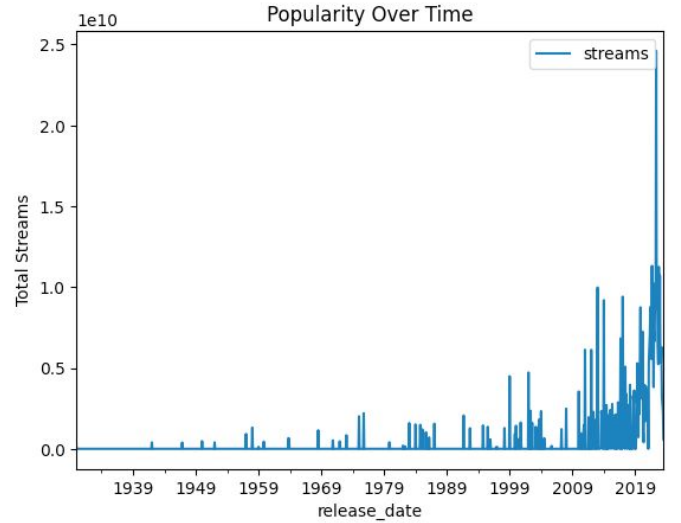

Fig. 2. Platforms Charts

Fig. 3. Stream Trend

C. Artist Impact

Analyze how artist involvement and attributes relate to a song's success. Visualization Suggestions:

- **Bubble Charts:** Map the success of songs (using size for popularity metrics like streams) against attributes like artist's genre or years active Fig[4].
- **Network Diagrams:** Show collaborations between artists and how these correlate with song success.
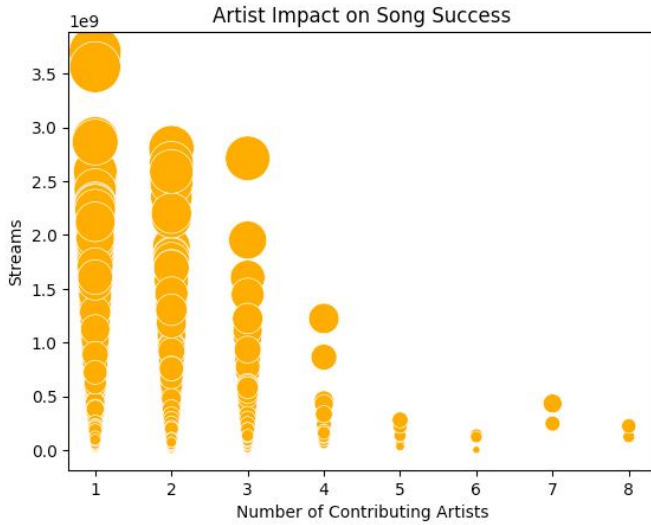
Fig. 4. Artist Impact

- **Bar Charts:** Aggregate song success metrics by artist or producer.

D. *Temporal Trends*

Identify shifts in music attributes and preferences over the year. Visualization Suggestions:

- **Line Graphs:** Trend lines for audio features or song popularity over the years.
- **Heat Maps:** Year-wise distribution of key features to show evolution over time.
- **Timeline:** Interactive timeline to explore key events in music (e.g., album releases, awards).

We will be adding interactive filters by genre, artist, year, and platform to allow users to customize the data they see. We will also make sure that the visualizations are organized logically and allow users to navigate easily between different views.

V. RELATED WORK

While previous studies, such as Rohit Jayakumar Nair's "Data Wrangling - Spotify Dataset Analysis," [3] [4] have explored the relationships between song features and their popularity to enhance user experience and marketing strategies on Spotify, our project aims to advance this analysis by incorporating real-time data through the Spotify API. Nair's project focused on static dataset analysis to predict track popularity and user clustering for personalized song recommendations. In contrast, our approach will leverage the Spotify API to access live data, enabling us to perform dynamic visualizations of music trends as they evolve. This capability will allow us to not only analyze but also visualize the fluctuating popularity of artists and tracks in real time, offering a more immediate and actionable insight into music consumption patterns. Furthermore, by dynamically categorizing songs and artists based on real-time data, we can provide a more nuanced

understanding of genre trends and artist impact, enhancing predictive models and marketing strategies with up-to-the-minute accuracy. This method represents a significant shift from static to dynamic analysis, offering a more robust toolset for stakeholders to respond swiftly to market changes and listener preferences, thereby refining the user experience with unprecedented precision.

Our project uses advanced analytical techniques to analyze music consumption dynamics . By tracking changes in musical preferences and detecting seasonal trends, we aim to provide a comprehensive view of the digital music landscape, enhancing predictive accuracy and strategic decision-making for music industry stakeholders.

VI. SUMMARY

This project explores the digital music consumption landscape, focusing on major streaming services like Spotify, Apple Music, Deezer, and Shazam. The dataset from 2023 provides insights into track attributes, streaming statistics, and audio features, as well as the musicality of songs. The challenge lies in decoding the factors driving song popularity across platforms and understanding how these factors influence listener preferences over time. By integrating data from the Spotify API, the project aims to enrich the dataset with genre, artist influences, and real-time popularity metrics. This integration will enable sophisticated, multi-dimensional visual analyses, generating dynamic visualizations that reflect shifts in artist popularity and audience engagement. The project aims to provide a nuanced analysis of current music industry trends, providing industry stakeholders with valuable insights to forecast future movements, tailor marketing strategies, and enhance the overall listener experience. The exploration of this enriched dataset will uncover actionable insights that will deepen our understanding of digital music consumption dynamics and inform strategic decisions in the music industry.

REFERENCES

[1] "Top spotify songs 2023." https://www.kaggle.com/datasets/nelgiriyewit hana/top-spotify-songs-2023?resource=download.
[2] "Spotify api." https://developer.spotify.com/documentation/web-api.
[3] "Spotify data analysis." https://rstudio-pubs-static.s3.amazonaws.com/6 96094_3daa60f42bb24af4abaa44ce46e28ea0.html.
[4] "Spotify sample projects." https://www.simplilearn.com/tutorials/data-a nalytics-tutorial/spotify-data-analysis-project.