

Group 4: Project Midterm

Amartya Dutta, Erin Wienke, Nabayan Chaudhury, and Timothy Kennedy

April 14, 2024

1 Introduction

The goal of our project is to analyze and visualize the characteristics of the top “50 tracks on Spotify in 2023” a dataset available on Kaggle to uncover patterns and trends in music preferences. We aim to explore various dimensions such as genre diversity, artist representation, tempo, and mood across the most popular tracks. This project seeks to present these findings in an engaging, intuitive manner to highlight how these elements contribute to a song’s success and popularity. The aim would also be to highlight what properties contribute to making a certain song popular.

2 Project Modifications

The primary aim of developing an interactive dashboard remains unchanged. One thing we decided upon was, in addition to the existing “Spotify Top 50 Tracks 2023” we plan to incorporate another dataset called “Most Streamed Spotify Songs 2023”. This dataset will be used to integrate further metrics such as the number of streams and the number of playlists featuring the songs. Integrating the “Most Streamed Spotify Songs 2023” dataset with the original dataset might introduce complexities in aligning data points between the two sources. Differences in dataset structures, naming conventions, or data granularity can complicate the integration process. Ensuring that both datasets are consistent in quality and format is crucial. Discrepancies in data collection methods or time frames covered by each dataset can lead to inconsistencies that might skew analysis results. Deciding which features are relevant and avoiding redundancy becomes challenging with more features available. Furthermore, presenting additional data such as streaming counts and playlist appearances must be done in a way that enhances, rather than confuses, the user experience. With the addition of new data points, the complexity of the interactive dashboard increases. Designing a user interface that accommodates additional information without compromising usability and aesthetic appeal requires careful planning and testing. To effectively analyze the expanded dataset advanced statistical methods might be needed. This can help in accurately identifying patterns and relationships between the added features and song popularity. These

are some of the potential challenges we might face in the future and accordingly, we will have to tackle them.

3 Design

Our team's design charrette went extremely well. The team was able to develop ideas and finalize an overall design for the project. When performing the charrette, we ran into new ideas, particularly in different visualizations that we could use to showcase data. Our final design will be a dashboard that showcases many different sub-visualizations that relate to different aspects of the top 50 track on Spotify. Within this visualization, a user will be allowed to select a song to view its specific characteristics within each of the sub-visualizations.

- **Dashboard Layout:** The dashboard will display comprehensive information about each song using the available attributes from the datasets.
- **Interactive Selection:**
 - **Song and Artist Selection:** Users will be able to select songs or artists from a list. Selecting an item will highlight relevant sections of the dashboard that pertain to the chosen song or artist.
 - **Attribute Distribution Selection:** Users can select specific regions within attribute distributions (like danceability, energy levels, etc.). This selection will guide them to songs or artists that match the chosen attribute profiles.
- **Relationship Exploration:**
 - **Related Songs and Artists:** Upon selecting a specific song or artist, the dashboard will use the attribute values of the selected item to suggest related songs or artists. This could be based on similarity in musical attributes or popularity metrics.

4 Tasks

- **Domain Problem:** Music analysis requires understanding complex relationships between audio attributes, popularity, and platform presence of songs.
 - **Task:** Enable users to explore and analyze these relationships to discern patterns and trends.
 - **Design Component:** The dashboard integrates interactive charts and graphs that represent various attributes like danceability, energy, and platform presence. Selecting a song or artist highlights relevant data across these visualizations, making it easier to understand how individual songs or artists compare against broader trends.

- **Domain Problem:** Users need to identify songs or artists that fit specific musical or popularity criteria.
 - **Task:** Allow users to select specific attribute ranges or values to filter songs or artists that meet these criteria.
 - **Design Component:** Interactive sliders and selection tools on attribute distributions (e.g., valence, tempo, streams) enable users to define and adjust their criteria dynamically. The dashboard responds by listing songs or artists that fall within these selected attributes.
- **Domain Problem:** Discovering new music or artists based on preferred attributes or existing favorites is a common user goal.
 - **Task:** Provide recommendations of songs or artists based on selected attributes or similarities to chosen songs or artists.
 - **Design Component:** Upon selecting a song or artist, the system calculates similarity in attributes and displays related songs or artists. This feature uses a similarity algorithm that considers multiple attributes, enhancing user discovery and exploration.
- **Domain Problem:** Users often wish to compare the popularity and presence of songs across different music platforms to gauge broad appeal.
 - **Task:** Facilitate cross-platform comparison of songs to highlight their popularity and presence on different streaming services.
 - **Design Component:** The dashboard provides comparative visualizations that depict each song’s ranking and presence on Spotify, Apple Music, Deezer, and Shazam. The selection of a song updates these visualizations to show detailed cross-platform data, enabling direct comparison.

5 Measuring Progress

The midterm checkpoint that we had set was as follows:

- **Data Pre-processing and Correlation Analysis** Acquire the "Top 50 Tracks datasets in 2023" from Spotify on Kaggle and conduct initial data cleaning and pre-processing. Data cleaning including handling missing values and standardizing formats is to be performed. Exploratory data analysis to understand dataset characteristics will be conducted. Basic statistics and trends are identified. Furthermore, we will analyze the correlation between musical features and track popularity. A preliminary report with visualizations and findings is prepared.

Accordingly, we have completed both data cleaning and pre-processing of the data. We have also performed feature correlations between some of the features in the dataset and have generated visualizations for them. Moving on, we will

incorporate the new dataset and generate some more of these data visualizations and also start building the dashboard which will allow the user to view these in an interactive manner.

6 Challenges

One major challenge that the team has faced during the development of our visualization is scheduling. We all are extremely busy with our other obligations, including research and classwork, so at times it can be difficult to all schedule a time to meet to discuss our work. One way we have attempted to combat this is by scheduling our meetings when the assignment is first released, as it gives us a longer period to attempt to have a meeting. Additionally, we have attempted to meet in smaller groups at a time, and then have the people who were able to attend those smaller meetings relay it in other small meetings to the rest of the group. Although this will continue to be a challenge in the development of our final visualization, we plan to continue to work as a team to mitigate the issues this may cause.