CS5764: Final Report

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

May 1, 2024

1 Introduction

The evolution of the music industry has been significantly impacted by technology, which has reshaped how music is produced, distributed, and consumed. However, there is a noticeable gap in the utilization and interpretation of streaming data for understanding music trends and preferences. Although streaming platforms like Spotify provide a wealth of data, there is a lack of comprehensive analysis that considers various musical elements such as genre diversity, tempo, and mood to understand what influences a song's popularity. This deficiency hinders the development of a detailed understanding of current music trends and listener preferences.

2 Motivation

The motivation for this project is derived from the opportunity to analyze vast amounts of music streaming data to unveil the components that contribute to a song's success. In the digital era, understanding these dynamics is essential for artists, producers, and industry stakeholders to align their outputs with listener preferences. [1] already deals with the management of songs audio features from a statistical point of view. In particular, it explores the data catching mechanisms enabled by Spotify Web API, and suggests statistical tools for the analysis of these data. By understanding the music preferences in the top 50 songs on Spotify in 2023, artists can develop a better understanding of how they can best tailor their music to reach the largest audience. Additionally, it can enable radio stations, festivals, recording studios, and producers to have a better understanding of what specifically they should invest resources into promoting [5]. This project aims to bridge the gap between raw data and actionable insights by examining the attributes of the top tracks on Spotify, thereby offering a deeper understanding of the contemporary musical landscape and enhancing the music listening experience. By employing advanced data analysis and visualization techniques, this research seeks to provide a comprehensive overview of the prevailing music trends and factors contributing to song popularity.

3 Problem Definition

The goal of this 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.

4 Related Work

Music is the most accessible and widely popular form of artistic expression, making it vital for us to understand its trends and their significance going into the future. Many statistical analyses have been conducted on musical trends. The work by [9] generally focuses on the qualitative attributes such as verbiage of musical pieces. While powerful, this doesn't provide a analysis of songs in their purest form - their generalized attributes. The closest to are paper to ours appears to be [19], but is in a area of time and a larger data set. The exploration of musical trends and characteristics, particularly through streaming platforms like Spotify, has garnered considerable attention in recent academic research. Spotify has published genre statistics, primarily with how popular certain genres are as well as the percentage increase of certain genres among certain years [2]. It would be interesting to see if the most popular genres contained the most popular songs, or if these songs were considered different genres. A study by [7] delves into the impact of technology on music streaming services, highlighting the role of digital marketing, social media, and the characteristics that contribute to Spotify's success and challenges. Additionally, [8] investigates the global music streaming trends to understand how local and global music preferences have evolved. This study by Spotify Research finds that preferences for local content have increased over time, and listener consumption is increasingly shaped by factors such as common official language and geographic proximity between countries. The digitization of music and its effects on global music markets suggest a trend towards Cultural Divergence, with increased diversity in music consumption across countries [12]. This analysis provides a broader context for our project, underlining the significance of exploring diverse musical elements and trends on platforms like Spotify to understand global listener preferences. The deeper interails of the system seen in [20] will be important in a graph we create in it's structure. Exploring simple yet elegant visual solutions for non-expert users in various domains, including music streaming data, underscores the potential of our project to make complex data accessible. By employing intuitive dashboards and interactive designs, we align with efforts to democratize data analysis, enabling users to discover insights into music preferences with ease [15].

In [3], the top 10 artists in terms of streams were examined. Additionally, the number of streams per day was also graphed. This gives insight into how users are using Spotify and listening to popular artists within the app. However, this differs from our approach as we are looking at the top 50 songs from 2023, we will not be examining streams, but instead looking at what makes these songs popular. Other studies go even further into user interactions with the services with how they interact with it, this paper does so with a massive data set [b], similarly Anderson's paper focuses on their interactions with Spotify's algorithm these approach are on focused on on the individual instead of what makes music popular [16]. Investigating the role of visualization in music genre exploration, a study contrasted bar chart and contour plot visualizations alongside mood control for user experience in discovering new music genres [11]. This approach highlights the importance of interactive elements and intuitive visualizations in enhancing user engagement and understanding, a principle that aligns with our project's goal to leverage advanced visualization techniques for Spotify data analysis. The use of Dash, Plotly, and Python in creating interactive visualizations for Spotify songs demonstrates the effectiveness of combining multiple diagrams for a comprehensive analysis. This approach informs our methodology, highlighting the importance of multidimensional views in understanding music trends and listener preferences [15].

5 Design

There are many music streaming platforms but Spotify has dominated the market over the last few years and currently is the largest one with millions of users. Spotify collects huge amounts of data from its users which presents a great amount of research opportunities in the data science community. Currently, music trend analysis often focuses on qualitative reviews or numerical charts without deep data-driven insights into the underlying characteristics of the tracks. While streaming platforms provide raw data on track popularity, there's a gap in comprehensive visualization that integrates various music features to offer a holistic view of what makes a track popular. Most analyses don't provide interactive exploration tools for the audience to engage with the data. [10] implemented an interactive dashboard for Spotify users with any level of expertise to perform exploratory, consumption, and analysis tasks on their personal Spotify streaming data. Our method draws inspiration from the Internet of Musical Things (IoMusT) and its focus on music data visualization. By incorporating visualization techniques that capture the essence of musical elements and styles, we aim to offer a multifaceted analysis of Spotify's top tracks, similar to how IoMusT projects facilitate the exploration of music through innovative interfaces [13]. Incorporating mood control via sliders for energy and valence into our visualization approach could significantly enhance user interaction and personalization. This method, proven effective in facilitating music genre exploration, can be adapted to our Spotify data analysis project to offer users a more tailored and engaging experience [11].

The approach we will follow in this project differs from the existing ones in a few ways. One of the most important being, the dataset is relatively new, so not a lot of in-depth exploration has been done using it. Furthermore, drawing inspiration from [10] we want to develop a dashboard which allows users to explore the Spotify Top 50 Tracks dataset through multiple lenses. By correlating these features with popularity, we can offer insights into the anatomy of a hit song in 2023. Music and its reasons for popularity can vary very quickly with time, hence making it important to perform analysis. The key aspects of our approach include:

- **Data Processing**: The first step involves collecting the Top 50 Tracks dataset from Spotify from Kaggle. This data will include various attributes of the tracks such as genre, artist, tempo, dance-ability, energy, and more. Data pre-processing will be essential to ensure the data is clean, structured, and ready for analysis. This includes handling missing values, standardizing formats, and extracting necessary features for the visualizations.
- **Correlation Analysis**: Apart from individual features, we want to visualize the correlation between different variables, such as the relationship between dance-ability and song popularity.
- **Interactive Visualizations**: Develop an interactive dashboard that allow users to explore the dataset. The dashboard will allow users to filter and sort data based on various parameters such as genre, release date, and audio features.
- **Insightful Metrics and Summaries**: The platform will also feature sections for data-driven insights and summaries, such as "Most Popular Genres of 2023" or "Top Artists of the Year."
- **Extend the data beyond 2023**: Additional data on listener demographics or external factors influencing track popularity could enhance the analysis. This supplementary data might need to be collected or inferred through additional research or public datasets. If additional data is available it could be extended beyond just the year 2023 to previous years to see if what makes a song popular has changed over time.

6 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.

7 Development

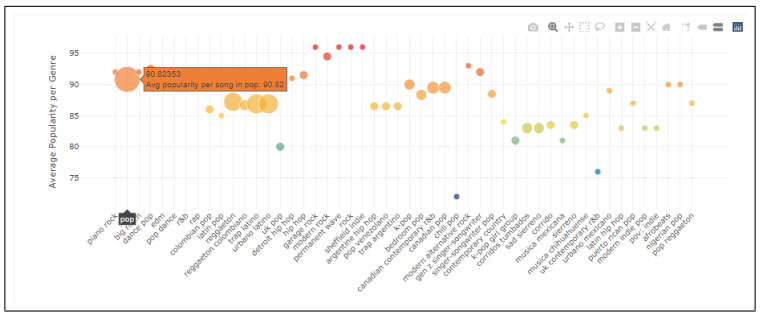
7.1 Developing the dataset

We used the 'Spotify Top 50 Tracks 2023' data set and merged it with the 'Most Streamed Spotify Songs 2023'. The 'Spotify Top 50 Tracks 2023' has 50 data points and 19 feature values for each. Meanwhile, the 'Most Streamed Spotify Songs 2023' has 953 data points and 24 feature values. The aim is to fetch some additional feature values for the top 50 songs from the second dataset. An interesting observation was that even though the latter data set had 953 data

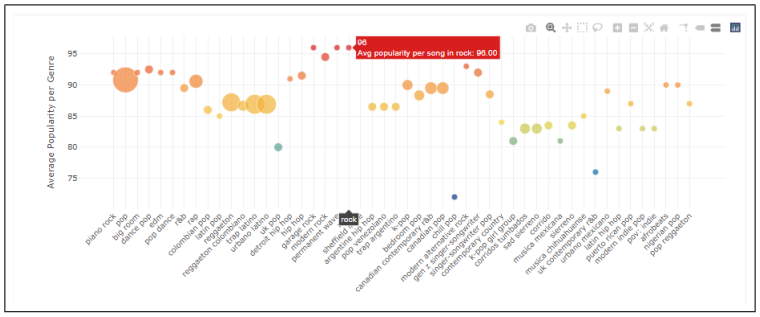
points, some songs in the top 50 were not present in it. There were 5 songs in the Top 50 that were not present in the Most Streamed dataset.
We are specifically interested in fetching the following additional feature values: Artist count for that song, Number of Spotify playlists the song is included in and Total number of streams on Spotify.
After merging the features for the top 50 songs, we checked for presence of NaN values in the additional features. We noticed there were 5 rows with NaN values for only the new features. Additionally, we notice that there are more than 50 data points after merging. Upon further investigation, we notice that 3 songs were repeated twice. Apparently the same Songs are present in the 'Most Streamed Spotify Songs 2023' more than once. We also noticed that the year of release varied for the same song in the 'Most Streamed Spotify Songs 2023'. Thus we decide to also include the Year when the song was released feature and then dropped the repeated songs with an older year of release.
We tried to replace the NaN values for the artist_count,in_spotify_playlists and streams. While we succeeded in replacing the NaN values for the artists by fetching additional artist names present in the Song title. The song titles would have key-words like 'feat.' and 'with' which would imply more artists and we used this information to replace the NaN values for artist count. In order to replace the NaN for the streams features, we calculate the Pearson's correlation coefficient between 'popularity' and 'streams' for the non-NaN rows. However, the coefficient value was only **0.18**. This implied that there was barely any correlation and so we will not be able to infer the stream values using the popularity feature. We face a similar challenge for 'in_spotify_playlists', thus we decided to drop these two features and proceed with only artist count as the additional feature.

7.2 Developing the dashboard

The dashboard was developed to provide an interactive, data-driven exploration of musical trends and metrics through a variety of dynamic visualizations. Central to this development was the transformation of prepared data into a cohesive dashboard comprising HTML, JavaScript, and CSS elements. Utilizing libraries such as D3.js and Plotly.js, the team constructed detailed visualizations that encapsulate song popularity, artist performance, genre characteristics, and more.
The process began with the preprocessing of Spotify's song and artist data, converting JSON data into visually interpretable formats. This involved aggregating song data by attributes like danceability, energy, and valence, and presenting this in various chart forms—bar plots, radial plots, and scatter plots. Each type of chart was tailored to convey specific aspects of the data: for instance, bar plots for displaying the number of songs per genre, scatter plots for artist popularity, and radial plots for song attributes.
To facilitate user interaction, navigation tabs were implemented, allowing users to seamlessly switch between different data views, such as songs, genres, and artists. The user interface was designed with responsiveness in mind, ensuring accessibility across various devices. Screenshots and interactive elements were



(a) Average popularity of Pop songs



(b) Average popularity of Rock songs

Figure 1: Average popularity over songs from each genre visualized

incorporated to enhance user engagement and to provide real-time insights into the data.

Users can employ this dashboard to discover the popularity of tracks, explore the attributes of top genres, and analyze the impact of artist collaborations on song popularity. Through interactive filters and dynamic visualizations, users can delve into the specifics of the music industry's current landscape, uncovering detailed insights into what makes a song or artist stand out in 2023. This dashboard serves as a powerful tool for industry stakeholders, music enthusiasts, and researchers alike, providing a comprehensive understanding of music trends and their implications.

8 Results

Our visualizations offer an in-depth analysis of song popularity, genre characteristics, and artist performance for 2023, providing valuable insights into current music industry trends. Analysis of the top 50 songs on Spotify reveals a relatively uniform popularity distribution, indicating a balanced listener preference across these tracks. This uniformity suggests that there is stiff competition among the top tracks, with no single song dominating the popularity charts.

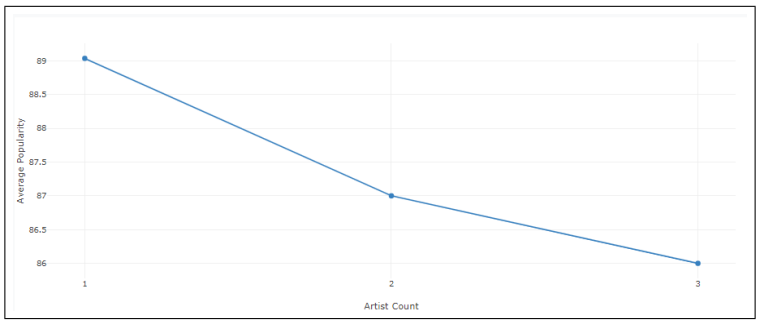


Figure 2: Impact of artist collaboration visualized

Additionally, the radar chart detailing song attributes highlights a trend towards moderate energy and danceability, with lower levels of acousticness, liveness, and speechiness. This pattern likely points to a preference for electronic or studio-produced music over live performances or spoken word content.
Genre-specific data shows a predominance of pop and its sub-genres, such as dance pop and Latin pop, which align with the global trend of pop music's widespread appeal. Despite the lower song counts in genres like jazz and classical, their popularity remains comparably high, indicating a dedicated, though possibly niche, listener base (Figure 1). The visualization of average attributes by genre further delineates the unique profiles of different music styles, for instance, EDM displaying high energy and danceability, while R&B features higher valence, suggesting more emotionally positive content.
The scatter plot examining the popularity of tracks by specific artists underscores strong brand or artist loyalty, as certain artists consistently produce highly popular tracks. Moreover, the chart titled "Does Collaboration Help?" presents a notable trend where the average popularity of songs decreases as the number of collaborating artists increases (Figure 2). This could suggest that while collaborations are common, they do not necessarily lead to higher song popularity and might result in a dilution of the artists' individual styles.
These comprehensive insights are crucial for artists, producers, and record labels, providing a data-driven foundation to tailor content, predict future hits, and strategize album releases and marketing campaigns effectively. This analytic approach underscores the importance of understanding listener preferences and market trends in shaping the future of music production and distribution.

9 Discussion

Overall, the development of this dashboard aided in the discovery of trends in the top 50 songs on Spotify in 2023. By understanding these trends, artists, labels, and producers can have a better understanding of what music actually sells, enabling them to pursue ventures that would fit into these categories. Additionally, a better allocation of resources will occur. Outside of just those

making the music, the understanding of trends of the top 50 songs on Spotify will allow artifical intelligence and algorithms to make better predictions and suggestions within the app. This will allow for users to be more satisfied with their experience using the app, causing them to be less likely to seek out and use competitors.
In the introduction, the goal of understanding and analyzing various musical elements such as genre diversity, tempo, and mood to understand a song's popularity and therefore music trends and preferences was provided. Using this as the target for the visualization, we can say that our visualization meets the challenges set forth in the introduction. The sections within the dashboard allow a user to analyze each of these specific elements individually in relation to the top 50 songs on Spotify from 2023, enabling a user to have a better understanding of what makes a hit song.
This motivates future work as it allows a viewer to determine whether there is any relation between specific elements of a song and whether that makes the song popular. Additionally, past years can be considered to have a stronger understanding of the trends among the hit songs of each year. By doing this, researchers can potentially predict the popularity of songs, as well as determine what the charts might look like in a given year. By doing this, artists, labels, and producers can try and tailor their sounds and the content they generate to attract the largest audience.

10 Final Progress

In the original report as well as the midterm report, the proposed final "exam" to check for success was the deployment of an interactive dashboard for data exploration. This dashboard would have the ability to filter and sort data based on various attributes. Additionally, the data would be broken into sections that allowed for a user to better understand the data with analytics to support the insights contained within the visualizations. Summaries should also be presented with each of the sections and visualizations.
Based on the proposed final assessment, our team has succeeded in meeting our goal for the semester. If we had done anything differently, we would've wished that we could have added additional years to our dashboard rather than just 2023. This would allow for an expansion of the information that we presented, as well as allowed us to view additional trends and see if the original trends we found hold from year to year.

11 Attribution

For the initial report, the team evenly split up the research and writing that was conducted. The same process took place for the design charrette. As for the midterm report, Erin and Amartya focused on the write-up and developing an outline for the dashboard, while Nabayan and Timothy focused on the

data pre-processing. For the final assessment, Erin focused on finalizing the outline for the dashboard as well as worked on the final write-up. Amartya and Nabayan contributed heavily to the actual design and completion of the dashboard. Timothy completed the finishing touches on the dashboard before recording the demonstration to showcase the final work.

12 References

1. Sciandra, M. and Spera, I.C., 2022. A model-based approach to Spotify data analysis: a Beta GLMM. Journal of Applied Statistics, 49(1), pp.214-229.
2. Lindner, J. (2023, December 16). Must-know Spotify genre statistics [latest report] • Gitnux. GITNUX. https://gitnux.org/spotify-genre-statistics/
3. Oraji, A. (2023, November 14). Detailed Spotify Data Visualization Analysis Report. Medium. https://medium.com/@ali.oraji/detailed-spotify-data-visualization-analysis-report-49caf27591d9
4. Spotify. (2022, November 30). Learn about those music genres you may not have heard of. For the Record. https://newsroom.spotify.com/2022-11-30/learn-about-those-music-genres-you-may-not-have-heard-of/
5. Zelenova, O. (2024, March 15). Why Spotify's music recommendations are so accurate. Kill the DJ. https://killthedj.com/why-spotify-recommendations-are-so-spot-on/
6. Corinthios, A. (2024, March 19). The biggest takeaways from Spotify's annual Music Economics Report. Spotify. https://newsroom.spotify.com/2024-03-19/loud-clear-music-streaming-royalty-data-artist-payments/
7. Li, J. (2022). Analysis of The Trend of Spotify. BCP Business & Management, 34, 919-926. https://doi.org/10.54691/bcpbm.v34i.3112
8. Way, S.F., Garcia-Gathright, J. and Cramer, H., 2020, May. Local trends in global music streaming. In Proceedings of the international AAAI conference on web and social media (Vol. 14, pp. 705-714).
9. Menten, M., Ng, K., O'Rourke, T. and Holmes, B., Temporal Trends in Music Popularity.
10. Ivanova, I. and Engstad, J.S., Explorify: A Personalized Interactive Visualization Tool for Spotify Listening History.
11. Yu Liang and Martijn C. Willemsen. 2021. Interactive Music Genre Exploration with Visualization and Mood Control. In 26th International Conference on Intelligent User Interfaces (IUI '21). Association for Computing Machinery, New York, NY, USA, 175–185. https://doi.org/10.1145/3397481.3450700
12. Bello, P., Garcia, D. Cultural Divergence in popular music: the increasing diversity of music consumption on Spotify across countries. Humanit Soc Sci Commun 8, 182 (2021). https://doi.org/10.1057/s41599-021-00855-1
13. P. Lončar, "Internet of Musical Things and Music Data Visualization," 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2022, pp. 1501-1506, doi: 10.23919/MIPRO55190.2022.9803404.
14. Espejo, A. (2023). Soundscape - Interactive Data Visualisations of Spo-

tify's Top Songs. Wellington Faculty of Engineering Symposium. Retrieved from https://ojs.victoria.ac.nz/wfes/article/view/8399
15. Clavadetscher, Sarah & Schlotter, Michael & Christen, Nadine & Streitberg, Juliane & Burch, Michael. (2024). Dashboard Design: Interactive and Visual Exploration of Spotify Songs. 10.5220/0012359100003660.
16. Anderson A, Maystre L, Anderson I, Mehrotra R, Lalmas M. Algorithmic Effects on the Diversity of Consumption on Spotify. In: Proceedings of The Web Conference 2020. WWW '20. Association for Computing Machinery; 2020:2155-2165. doi:10.1145/3366423.3380281
17. Zhang B, Kreitz G, Isaksson M, et al. Understanding user behavior in Spotify. In: 2013 Proceedings IEEE INFOCOM. ; 2013:220-224. doi:10.1109/INFCOM.2013.6566767
18. Zhang B, Kreitz G, Isaksson M, et al. Understanding user behavior in Spotify. In: 2013 Proceedings IEEE INFOCOM. ; 2013:220-224. doi:10.1109/INFCOM.2013.6566767
19. Espejo A. Soundscape - Interactive Data Visualisations of Spotify's Top Songs. Wellington Faculty of Engineering Symposium. Published online October 2023. https://ojs.victoria.ac.nz/wfes/article/view/8399
20. South T, Roughan M, Mitchell L. Popularity and centrality in Spotify networks: critical transitions in eigenvector centrality. Journal of Complex Networks. 03 2021;8(6):cnaa050. doi:10.1093/comnet/cnaa050