CS 5764 Project Proposal: Visualizing the Spotify Soundscape

Amartya Dutta
PID : amartya

Erin Wienke
PID : erinw01

Nabayan Chaudhury
PID : nabayanc

Timothy Kennedy
PID : ktimothy20

1   Introduction

1.1   Problem

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.

The goal 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.

1.2   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.

2 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].

3 Approach

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.

4 Impact

The impact of this project on the visualization of the Spotify Top 50 Tracks dataset in 2023 extends across various sectors. By offering interactive explorations through genres, artists, tempo, and audio features, it aims to illuminate the elements contributing to a song's success in 2023. As the number of independent artists continues to grow (accounting for half of what the industry generated on Spotify), it is becoming ever more important for artists to understand what defines popular music [6]. This analysis benefits artists and producers by highlighting current trends and listener preferences, guiding their creative processes towards potentially more successful outcomes. The shifting dynamics in global music consumption, propelled by digitization, inform our project's potential impact on understanding cultural preferences. By analyzing Spotify's top tracks, we contribute to the discourse on music globalization, offering insights into how digitization shapes listener diversity and preferences worldwide [12].

Music industry professionals, from analysts to executives, could leverage the dashboard to refine marketing and development strategies, tailoring their efforts to align with evolving music tastes. Reflecting on a project that developed an interactive web application for visualizing Spotify's music ranking data, we foresee our project similarly empowering users with the tools to uncover insights into music trends and preferences. By enhancing user interaction with data through effective visualization, we contribute to informed decision-making and a deeper understanding of the music industry [14].Academics and students in music and cultural studies will find a rich resource for understanding and teaching contemporary music dynamics. The general public and music enthusiasts gain a deeper appreciation and understanding of their favorite songs, democratizing music analytics previously available to industry insiders. Lastly, streaming services could improve recommendation algorithms and user engagement based on insights from this project, enhancing the listening experience for

millions. Overall, this initiative promises to deepen our understanding of musical preferences, foster informed creative processes, and improve music consumption experiences.

5 Checkpoints

In order to ensure that the project progresses smoothly and meets its objectives, the following checkpoints serve as markers to assess the project's progress:

- **Checkpoint 1: 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 correlation between musical features and track popularity. A preliminary report with visualizations and findings is prepared.

- **Checkpoint 2: Development of Interactive Visualizations and Dashboard** Develop and deploy an interactive dashboard for data exploration. Functional dashboard allowing data filtering and sorting based on various attributes is deployed. Usability tests confirm the interface's effectiveness. Create sections for data-driven insights and summaries within the dashboard. Sections like "Most Popular Genres of 2023" are included with data analysis supporting these insights. Possibly extend analysis beyond 2023 and integrate additional data sources if available. Updated analysis will reflect broader scope and additional factors. Provide final report will all analytical findings along with an interactive dashboard.

6 Challenges and Outcomes

Throughout the course of this project, we anticipate facing several challenges. Accurately correlating musical features with track popularity involves complex statistical analysis, which may be challenging due to the multi-dimensional nature of music and its subjective appreciation. Another potential challenge is the development of interactive visualizations and a user-friendly dashboard. Ensuring that the dashboard is intuitive for users with different levels of expertise requires careful design and testing. Additionally, extending the dataset beyond 2023 and integrating supplementary data could face hurdles related to data availability, compatibility, and privacy concerns. Additionally, another challenge that could arise is that Spotify has 1742 genres classified [4]. There is a possibility that each song within the Top 50 dataset is classified as a different genre. It may also be hard to determine which of these genres can be grouped if the need arises, as some of them are extremely niche.

Despite these challenges, the expected outcomes of the project are substantial. By successfully navigating data processing and analysis, we aim to provide a comprehensive and intuitive platform that sheds light on the dynamics of music popularity. The interactive dashboard will enable users to explore music trends, discover relationships between different musical attributes, and gain insights into the characteristics of successful tracks. By highlighting trends and preferences in the music industry, the project has the potential to influence future music production and marketing strategies. Ultimately, the project seeks to enrich the understanding of music's evolving landscape, fostering a deeper connection between listeners and the music they enjoy.

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 Spotify'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