Exploring Musical Diversity on Spotify: An analysis of audio features across multiple artists

A. Kathiresan PL - *20BCE2400*, B. Annanya Mangla - *20BCE0700*, C. Sisira Reddy - *20BCE2448*, D. Ishika Garg - *20BCE2793*, E. Takshu Yadav - *20BCE2796*

School of Computer Science and Engineering

Vellore Institute of Technology, Vellore

Abstract -

Exploring Musical Diversity on Spotify: An Analysis of Audio Features Across Multiple Artists" is a comprehensive research project that delves into the intricate landscape of music on the popular streaming platform, Spotify. The study encompasses a multifaceted analysis, focusing on key aspects such as identifying top artists with high listenership, evaluating audio features for genre and artist differentiation, examining evolving personal listening habits, and assessing emotional sentiment within specific artists' song catalog. The project's findings not only spotlight prominent artists on Spotify but also shed light on their dedicated fan bases. Through the evaluation of audio features, the study provides profound insights into the defining characteristics of various musical genres and artists, enriching our understanding of the platform's diverse musical ecosystem. Additionally, personalized insights are gleaned from the analysis of individual listeners' Spotify data, revealing patterns in music consumption over time. The emotional content analysis adds a unique dimension, uncovering the mood and sentiment conveyed within an artist's discography. This project contributes significantly to a deeper comprehension of the expansive and diverse musical landscape on Spotify, benefiting both music enthusiasts and industry professionals alike.

Furthermore, the project contributes to our understanding of the evolving dynamics in contemporary music consumption. By scrutinizing personal listening habits over time, the research captures the dynamic interplay between users and the extensive musical library on Spotify, offering insights into the shifting preferences and patterns of today's listeners. The nuanced analysis of emotional content within artists' discographies not only enhances the individual user experience but also provides valuable information for artists and industry stakeholders. This project, through its holistic examination of audio features and emotional nuances, emerges as a pivotal exploration, bridging the realms of data-driven analysis and the emotive intricacies inherent in the realm of music. It not only informs our comprehension of the musical landscape on Spotify but also paves the way for future endeavors in understanding the symbiotic relationship between technology, user preferences, and the ever-evolving artistry of musical expression.

Index Terms - Analytics, Dashboard, Music Consumption Trends, Predictive Analysis, User Behavior.

I. Introduction

In the digital age, music streaming platforms like Spotify have revolutionized the way we consume and interact with music. "Exploring Musical Diversity on Spotify" is a multifaceted project that takes a deep dive into this transformative musical landscape. With data collection and interactive visualizations, the project seeks to uncover the richness and complexity of music on Spotify.

The project's scope includes identifying the top artists with the largest listener bases and most followers, allowing us to recognize the most influential figures in the industry. Analysis of audio features like tempo, danceability, and energy offers insights into the characteristics that define various music genres and artists.

Moreover, a personalized dimension is added by examining the listener's own Spotify data, revealing trends in listening habits over time. Additionally, the project assesses the emotional sentiment of songs by specific artists, adding a new layer of understanding to the musical experience.

This project caters to music enthusiasts, researchers, and industry professionals, offering a comprehensive view of the multifaceted world of music on Spotify.

II. Literature Survey

[1] Analysis of The Trend of Spotify

The research paper on Spotify's business model utilizes bar charts to visually convey key financial and user-related metrics. Figures depict Spotify's worldwide revenue growth from 2013 to 2021, monthly active users, ad-supported users vs. subscribers, revenue by segment, gross profit, gross margin, and free cash flow. The charts reveal a positive financial trend over the past seven years, with significant revenue growth. Monthly active users and ad-supported users show consistent increases, while gross profit and margin highlight segment variations. The paper emphasizes Spotify's strengths despite some financial fluctuations, positioning it as a distinguishable leader in the music streaming industry.

[2] What makes a Song Trend? Cluster Analysis of Musical Attributes for Spotify Top Trending Songs

Pie Chart :

The paper likely employs pie charts to visually represent the distribution of musical characteristics, such as genres, within the "Top 100 Trending Spotify Songs" dataset for 2017 and 2018. Pie charts offer a clear and concise way to illustrate the proportion of songs exhibiting specific attributes, providing valuable insights into the overall composition of the trending songs, enhancing the reader's understanding of categorical data distribution.

Scatter Plot :

The research utilizes scatter plots, such as the one in Figure 5, to explore the relationship between speechiness and loudness in the "Top 100 Trending Spotify Songs" from 2017 and 2018. Scatter plots are effective visual tools for revealing patterns or correlations between two continuous variables. In this context, the scatter plot offers a visual representation of how speechiness and loudness are related in the analyzed songs, contributing to a comprehensive understanding of musical characteristics and trends.

[3] Strategic marketing analysis : a case study of Spotify

Pie chart

The paper explores the global streaming music subscription market, noting Spotify's dominance, which has slightly decreased from 34% in 2019 to 33% in 2020. With an anticipated 14.7% CAGR between 2022 and 2030, major players like Spotify, Apple Music, Tencent Music, and Amazon Music vie for increased market share. The research

aims to analyze Spotify's market position, growth strategy, and challenges amidst intense competition. While specific details on pie chart usage are not provided, a pie chart might illustrate the proportional market share of key players, aiding in a clear and concise presentation of the competitive dynamics in the evolving music streaming industry.

[4] Using Visualizations to Encourage Blind-Spot Exploration

Scatter plot

The study employs scatterplots as a visualization tool, representing data with points determined by two variables. In Figure 2, bubble size signifies item set frequency, with larger bubbles indicating higher genre frequency. Colors distinguish genres, while bubble orientation conveys average artist hotness and whether the genre belongs to the global or user profile. The hover feature facilitates easy genre comparison between profiles, providing details on frequency, hotness, and genre name. Inference from the visualization reveals the user's preference for Pop as the most consumed genre, its global popularity, and the user's inclination towards more popular artists than the system's average user.

Bar - line chart

The study compares scatterplots with bar-line charts as visualizations. Bar-line charts combine bar and line charts, accommodating up to three variables. Bar charts were chosen as the base visualization for their proven effectiveness in recommender systems and their usage in platforms like MovieLens. In Figure 4, the chart displays item set names on the horizontal axis, genre frequency on the left-vertical axis (line chart), and average artist hotness on the right-vertical axis (bar chart). Unlike scatterplots, a single bar-line chart shows either user or global data, necessitating two separate charts for comparison. This approach prioritizes chart simplicity over simultaneous representation.

[5] Cultural Divergence in popular music: the increasing diversity of music consumption on Spotify across countries

Line chart

The study employs line charts for visualizations to depict trends over time in song, artist, and label diversity on platforms like Spotify and iTunes. Line charts effectively convey the evolving patterns, seasonality, and growth in diversity metrics, allowing a clear comparison between different elements. The choice of line charts is particularly suitable for presenting continuous data, aiding in illustrating the nuanced changes in the music charts, and facilitating a comprehensive analysis of the increasing diversity in songs, artists, and labels across the observed period.

[6] Explorify: A Personalized Interactive Visualization Tool for Spotify Listening History

Bar plot + Heatmap

The paper likely utilized a heatmap to effectively convey multifaceted information in a visually intuitive manner. Heatmaps excel in displaying diverse data sets simultaneously, ideal for showcasing three distinct views of the interactive barchart. Firstly, the heatmap accommodates the sorted barchart views by streaming time, artist popularity scores, and artist names side by side. This layout allows for quick comparisons between the different sorting methods, aiding in pattern recognition and insights.

Moreover, the interactive elements, such as the selection and hover actions, are efficiently showcased through color changes and pop-up displays. These interactions are easily integrated into a heatmap, enhancing user engagement and facilitating a deeper understanding of the data.

By employing a heatmap, the paper likely aimed to optimize the visual representation of complex information, offering a comprehensive yet accessible depiction of the diverse aspects of the Spotify analysis.

Dashboard creation

The paper likely utilized a dashboard for its Spotify analysis due to its ability to present multifaceted data in a coherent and user-friendly manner. Dashboards offer an integrated platform where diverse information can be visually represented and interactively explored. In this case, the interconnected views—like the 'Artist Genre Network,' 'Daily Listening Pattern,' 'Top 20 Artists,' and 'Detailed Listening History per Day'—create a comprehensive snapshot of the user's Spotify usage.

The dashboard's advantage lies in its capacity to allow users to navigate through different layers of data effortlessly. For instance, the primary view of the 'Artist Genre Network' likely provides an overview, while the supporting views offer deeper insights into daily listening habits or specific artist preferences. Additionally, the interactivity, where selecting a day in the 'Daily Listening Pattern' triggers the display of 'Detailed Listening History per Day,' enhances the user experience by enabling focused exploration.

Overall, the dashboard's design facilitates intuitive data exploration, enabling users to derive meaningful insights from complex Spotify usage patterns in a visually engaging manner.

[7] Effects of recommendations on the playlist creation behavior of user

Bar chart

Bar plots are an effective choice for illustrating distributions, especially for the distribution of selected topics by participants in your context. They offer a clear visual

representation of aggregated values across different contexts (as seen in Fig. 11a) and enable easy comparison between these categories.

The simplicity of bar plots allows for a straightforward understanding of the distribution patterns, making them ideal for conveying how topics are distributed among various contexts. Additionally, when detailed values for each topic within contexts and purposes are presented (as depicted in Fig. 11b), bar plots facilitate a concise yet comprehensive view of individual data points, aiding in identifying specific trends or disparities within the dataset.

[8] Promoting Music Exploration through Personalized Nudging in a Genre Exploration Recommender

Scatter plot

A scatter plot is a powerful visual tool chosen for depicting the relationship between Mean Squared Absolute Error (MSAE) and predicted consistency from a multilevel regression model due to its capacity to illustrate individual data points' distribution and the correlation between variables.

The choice of a scatter plot allows for the clear representation of multiple dimensions: MSAE on one axis, predicted consistency on the other. It enables the visualization of how closely the predicted consistency aligns with the actual MSAE values across different levels: track, artist, and tag.

Moreover, the inclusion of a 95% confidence interval in the shaded area adds a layer of statistical insight, offering a visual depiction of the uncertainty around the estimated regression line, thereby reinforcing the reliability of the model's predictions.

[9] Platforms, promotion, and product discovery: evidence from spotify playlists

Bar plot

Bar charts are an effective choice for visualizing the relationship between two categorical variables, such as the New Music Friday rank and Spotify chart appearance. This choice aligns with the need for a clear, comparative display of discrete data points.

The bar chart’s simplicity allows for easy interpretation of the New Music Friday ranks against the occurrence of appearances on Spotify charts. Each bar represents a distinct category (New Music Friday rank), and the height or length of the bar correlates directly with the frequency or significance of appearances on the Spotify chart.

Additionally, bar charts facilitate quick comparisons between different ranks and chart appearances, aiding in identifying trends or patterns. This visual form presents a snapshot of the relationship between the variables, enabling researchers to easily discern any potential correlations or dependencies between them.

[10] The audio features of sleep music: Universal and subgroup characteristics

Density plot

The utilization of density plots in depicting data distributions offers a nuanced perspective on the central tendencies and variability within different categories, such as Insight, Confidence, Essence, and Time. Unlike boxplots, density plots provide a continuous representation of the distribution, showcasing the entire range of values and emphasizing the shape and peaks within each category's data distribution. This method enables a more detailed understanding of the distribution's form, highlighting subtle variations that might not be as apparent in boxplots.

Furthermore, by employing density plots, the paper aims to present a more comprehensive view of participant responses across ICE-T dimensions. Unlike boxplots, which focus on summary statistics and outliers, density plots offer a visual narrative that allows for a deeper exploration of the data's distributional characteristics. This approach aligns with the need for a richer, more nuanced depiction of multi-category distributions, fostering a more intricate understanding of how scores vary across different dimensions and aiding in identifying intricate patterns or differences within the data. Ultimately, the choice of density plots supports a more detailed and insightful analysis of the ICE-T scores, enhancing the paper's ability to convey complex data relationships effectively within a succinct visual format.

[11] SpotiVis - Finding new ways of visualizing the spread of popular music

Line chart + map chart

The paper likely employed a line chart for depicting the music chart data due to its ability to illustrate temporal trends effectively. Line charts excel in showcasing changes over time, making them ideal for displaying how a song or artist's performance fluctuates across different national music charts over a specified period. The visual clarity of a line chart allows viewers to easily discern patterns, spot peaks or dips in popularity, and observe the overall trajectory of the song or artist's success.

In contrast, the geovisual representation on a world map with a colormap scale offers a spatial perspective, associating chart positions with countries. However, it might lack the

precision and immediacy in conveying temporal changes that a line chart provides. The line chart's linear progression chronicles the evolution of chart positions, facilitating a clearer understanding of the song or artist's performance dynamics over time. This choice likely aimed to prioritize conveying temporal trends comprehensively, making the line chart a suitable visualization method for the paper's analysis of music chart data.

Box plot

The use of boxplots allows for a clear understanding of central tendencies, such as median and quartiles, enabling easy comparisons between different categories (Insight, Confidence, Essence, and Time). Additionally, they efficiently display the variability within each category, showcasing the spread and presence of outliers, if any, without overwhelming the reader with excessive detail.

Moreover, by presenting multiple boxplots in Figures 8 to 10, the paper organizes and compares distributions across distinct categories, aiding in the identification of patterns or differences in participant responses. This approach facilitates a nuanced understanding of how scores vary across the ICE-T dimensions, providing a visual narrative that supports the paper's analysis and conclusions. Overall, the boxplot choice aligns with the need for a succinct yet informative depiction of multi-category distributions in the context of the ICE-T scores.

[12] What Kind of Music Do You Like? A Statistical Analysis of Music Genre Popularity Over Time

Line chart

The paper utilizes line charts to visualize the discrepancy between observed genre proportions and model-predicted probabilities within the Pazz & Jop model. Line charts effectively illustrate trends over 40 years, showcasing the divergence between critic opinions and model predictions, attributing it to the variability in critics' perspectives. They highlight the nuanced shifts in music trends, such as the rise of Rap/Hip-Hop in the late '80s and early '90s, Rock's gradual decline, the prominence of Folk among critics, and the evolving acceptance of Pop over time. Line charts offer a clear depiction of these long-term trends, smoothing out yearly fluctuations and emphasizing the overarching patterns within music criticism and consumer reception.

Box plot

The paper likely used boxplots to illustrate the forecasted genre proportions and their uncertainty due to multiple simulations. Boxplots effectively convey the distribution of simulated values by displaying the median (central line), quartiles (box edges), and potential outliers (whiskers). With 200 simulations for each predictor value, a boxplot efficiently showcases the range and variability of forecasted genre proportions. The

boxplot's visual simplicity allows for easy comparison across genres, highlighting potential variations and tendencies in forecasted proportions. Additionally, the use of boxplots aligns with statistical conventions, providing a clear summary of the distributional characteristics of the simulated data. This graphical representation aids in comprehending the uncertainty inherent in the forecasted values derived from ARIMA models, offering a visual depiction of the variability and range of potential outcomes for each genre's proportion.

[13] Music we move to: Spotify audio features and reasons for listening

Density plot,

The paper likely used density plots due to their ability to illustrate the distribution and overlap of data sets in a visually intuitive manner. Density plots provide a clear representation of the differences in Spotify audio features between baseline and dance music datasets by showcasing the concentration and spread of values for each feature. This visualization method allows for easy comparison, highlighting significant distinctions in feature distributions. It offers a comprehensive view of the statistical analysis results, emphasizing the varying characteristics between the two music categories, making it an effective choice for conveying nuanced differences succinctly and comprehensively.

Heatmap

A heatmap was employed in the paper to visually represent the intricate relationships between various musical characteristics of dance music and individuals' real-life experiences (RL). This graphical method efficiently showcases the distribution of participants' ratings for specific RL items across different subgroups of dance music. By displaying variations in music associated with distinct listening purposes, the heatmap reveals nuanced patterns, suggesting that certain subgroups of dance music might align more suitably with specific real-life experiences. Its use enables a comprehensive and easily interpretable presentation of how different musical traits correspond to diverse RL aspects, facilitating a detailed discussion in the paper's later sections.

[14] How playlist evaluation compares to track evaluations in music recommender systems

Scatter plot

A scatter plot serves as an ideal choice to portray the relationship between subjective user ratings and algorithm-predicted recommendation scores due to its ability to display individual data points while showcasing potential patterns or trends. The slight jittering of user ratings aids in visual clarity, preventing overlap. Additionally, incorporating a shaded area representing the 95% confidence interval adds statistical context, enhancing the plot's credibility by illustrating the range within which the true values are likely to fall.

[15] Exploring Music Rankings with Interactive Visualization

Tree Map

The Treemap effectively displays hierarchical data, representing the proportion of streams for various music genres within a specific week. Each rectangle corresponds to a genre, and its size indicates the stream volume. This visual approach aids in comparing and understanding the popularity of genres. The treemap, along with other visualization methods, aims to enhance user experience and engagement, providing a comprehensive way to explore and compare music rankings from sources like Billboard and Spotify.

III. Methodology

1. **Data Exploration and API Integration:**
   - **Exploratory Excitement:** The initial dive into the Spotify API opened a world of possibilities. The sheer volume of available data sparked excitement and set the stage for the project's data-driven narrative.
   - **Integration Challenges:** Navigating the intricacies of the Spotify API presented challenges. Overcoming these hurdles not only enhanced technical skills but also deepened the understanding of data retrieval mechanisms.

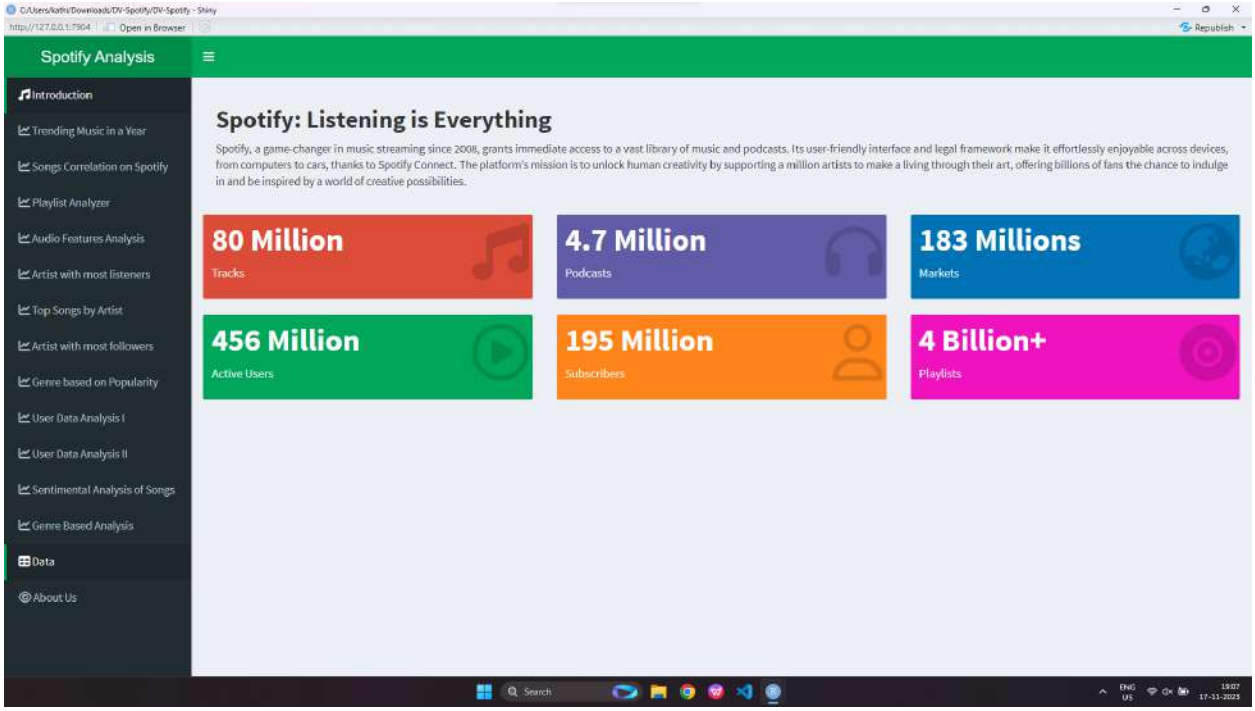
2. **Data Cleaning and Preparation:**
   - **Wrangling Narratives:** The process of cleaning and preparing the Spotify data was a journey of its own. Unexpected insights emerged during the handling of missing values and outliers, enriching the dataset and setting the foundation for meaningful visualizations.

3. **Exploratory Data Analysis (EDA):**
   - **Moments of Revelation:** EDA was a journey of discovery, uncovering patterns, trends, and relationships within the Spotify data. Each moment of revelation influenced the subsequent direction of the project, adding layers to the narrative.

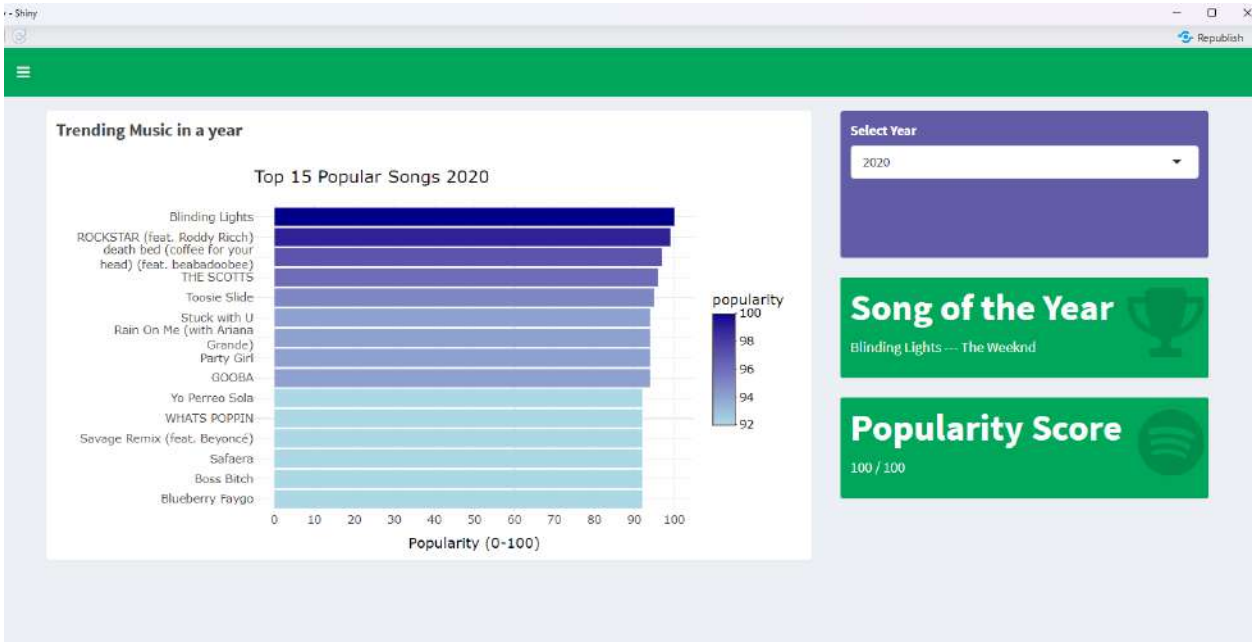
4. **Shiny Dashboard Creation:**
   - **Creative Decisions:** The design of the Shiny Dashboard was a canvas for creative expression. Layouts, tabs, and components were carefully crafted to provide an intuitive and aesthetically pleasing user experience.
   - **Vision for Interaction:** Interactive features were envisioned to empower users in exploring Spotify data dynamically. The dashboard aimed to be more than a visualization tool, offering an immersive journey through the musical landscape.
5. **Visualization Implementation:**
   - **Artistry with Plotly, ggplot2, and leaflet:** Utilizing Plotly, ggplot2, and leaflet was akin to painting with data. Each visualization was a piece of art, conveying the stories hidden within the Spotify dataset. The emphasis was on creating not just graphs but visual narratives.
   - **Reactive Programming Journey:** The use of reactive programming in Shiny brought the visualizations to life. The dynamic nature of the dashboard, responding to user input in real-time, added a layer of engagement that surpassed static visualizations.

IV. Results

Introduction Page:



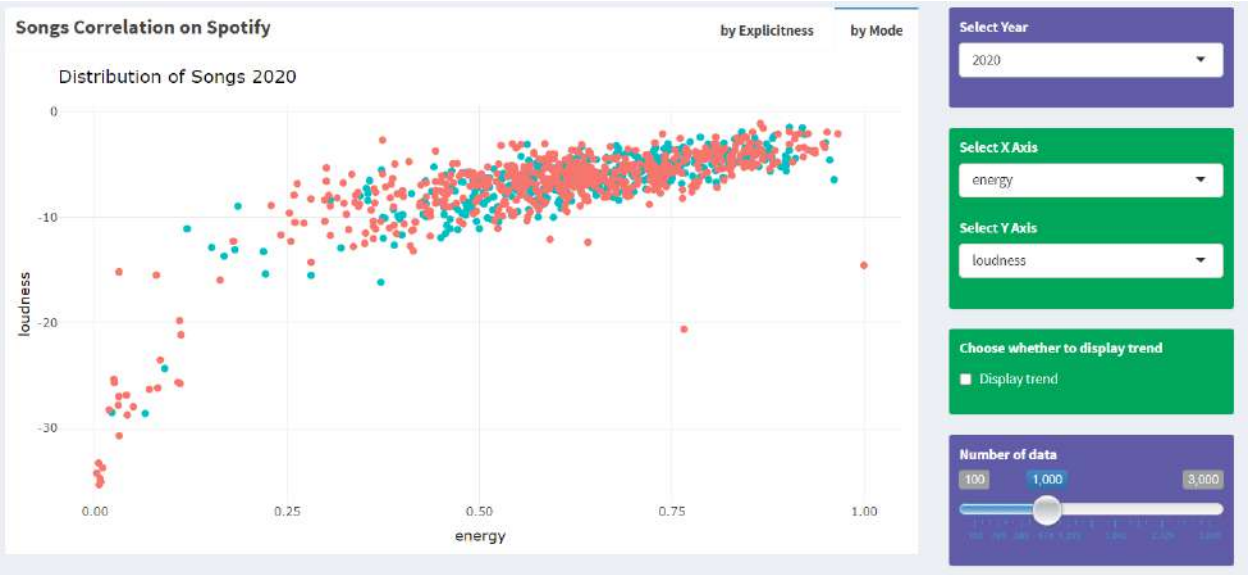
Trending Music in a year:



-> This plot displays us the top 15 popular songs of any year with the help of a horizontal bar chart.

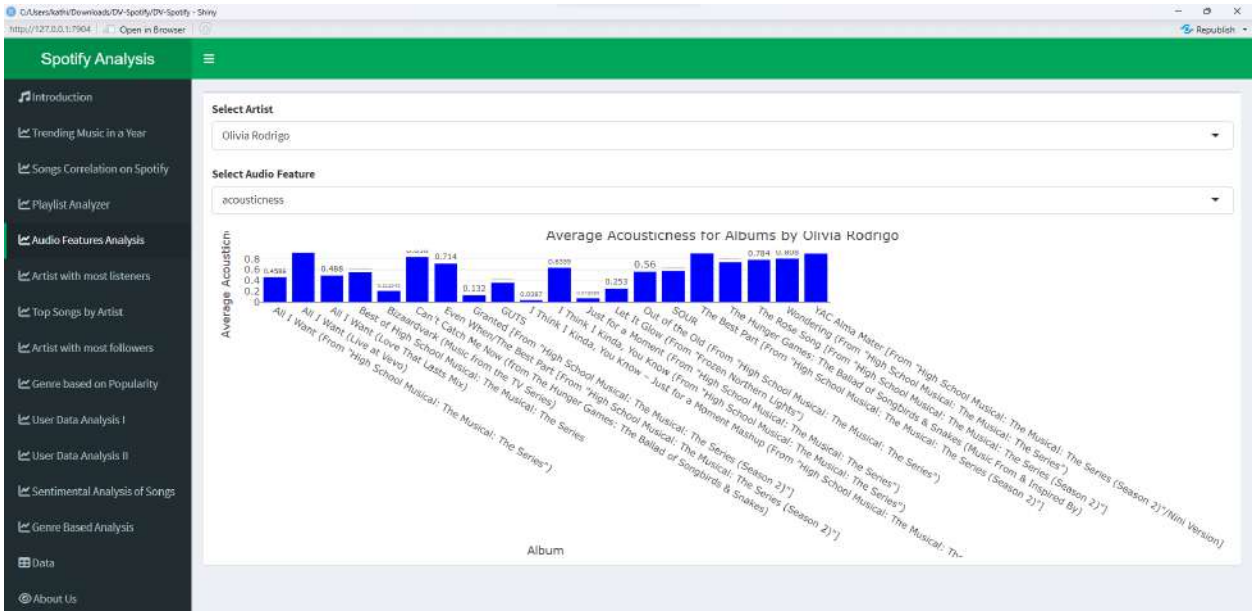
-> The interactivity in this plot is through the drop down of selecting the year, changing which plot shows the top 15 popular songs for that year based on their popularity score which is an attribute in the dataset.
->The song of the year is the 1st song in the 15 most popular songs and its score is also displayed.
-> Hovering over the bars shows us the exact popularity score of the songs.
-> This is a 2 hue color map used for popularity.
-> In 2020 blinding lights is the most listened to song with a popularity score of 100.

Songs Correlation on Spotify [2] :



->This is a scatter plot that allows users to explore and analyze correlations between different audio features of songs on Spotify for a selected year.
->In this plot we have interactivity in selecting the year, and the audio features for the x and y axes.
We can also decide if we want to display the trendline and the number of data points to be taken to make the plot.
-> There is an option to also switch between explicitness and mode which shows how they correlate with the audio feature.

Audio Feature Analysis [4] :

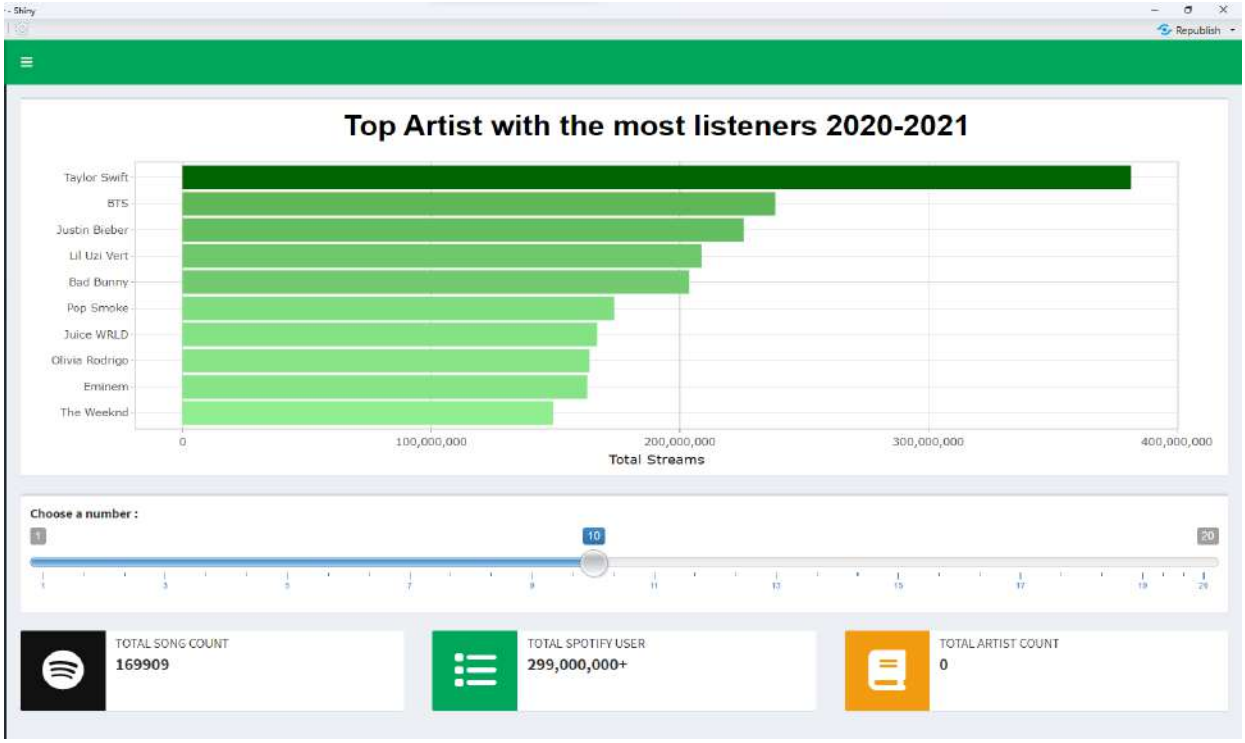


-> This is a vertical bar chart in which in which for the selected user's all albums the average of audio feature across is shown.

-> Here the user gets to choose the artist from the dropdown list which shows all the artists from our dataset. The user can also select the audio feature for which they want to see the average .

->In the above plot we can see that "All I want" and "YAC Alma Mater" have the highest average and "Just for a moment" the lowest.

Artist with most listeners:



-> This horizontal bar plot displays us the top artists with most listeners in the 2020-2021 time period.
-> We have an option to select the number of artists that have to be considered for this plot using a slider.
-> The color also represents the total streams here the darker the green the more streams.
-> For the above plot Taylor Swift has the most listeners in the time period with around 390,000,000 streams.

Top Songs by Artist:

-> This is a line chart that displays the top songs of an artist by taking the total number of streams of the songs.
-> Here the user gets to choose the artist from the dropdown list which shows all the artists from our datase.
-> The plot shows us that the most listened song under Eminem is Premonition with 10830400 streams.
-> Hovering over the green point in the line chart the user gets to see the exact number of streams.

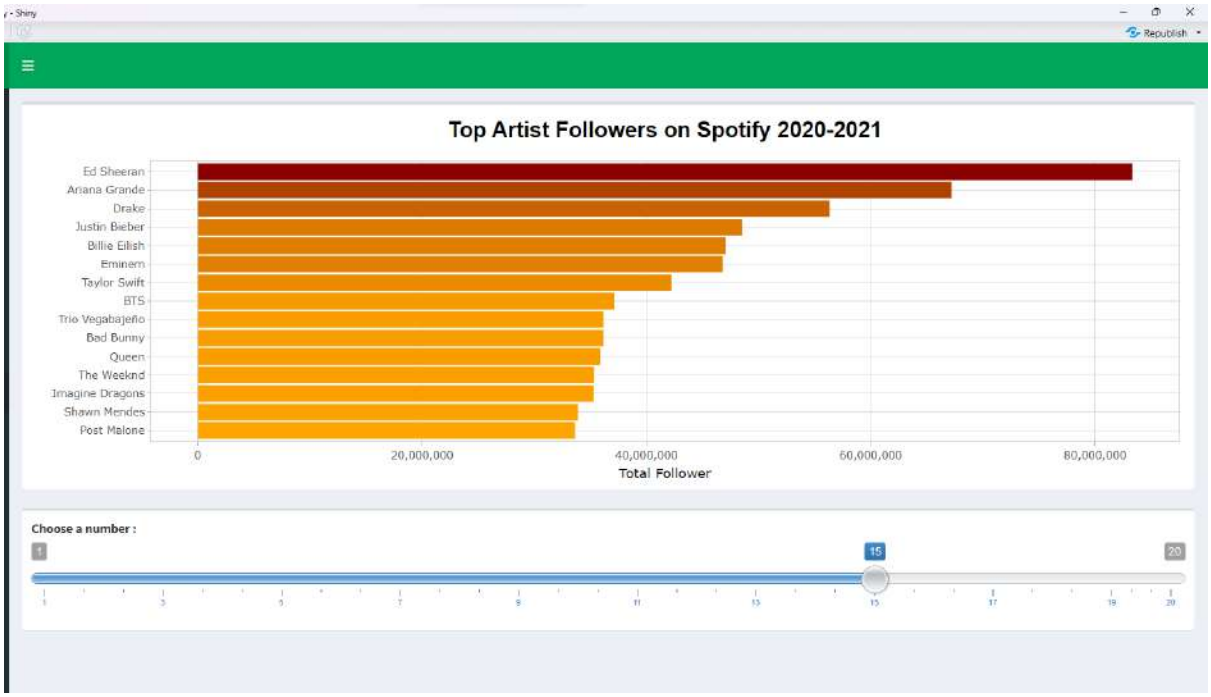
Spotify Playlist Visualizer:

-> This is a scatter plot allows the user to explore and analyze the relation between different audio features of songs under a playlist selected by the user.
-> The user can enter the playlist id which is at the end of the url, select the audio features for x and y axes and choose color if they want a 3rd variable to be consider and also select if they want to see a trendline.
-> According to the plot taken for example the lesser the popularity the lesser the danceability is visible in most of the cases. But this isn't a strong analysis as there are a lot of exceptions.

Genre Based Popularity:

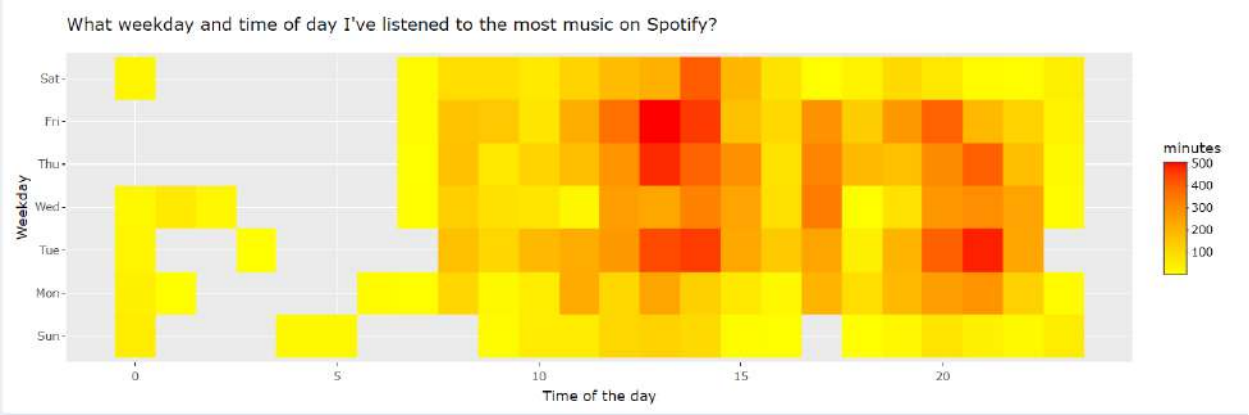
-> The word cloud is a visual representation of different genres of music based on their popularity on Spotify.
-> The genres are written in different colors and fonts to show their diversity and uniqueness.
-> The size of the genre indicates its popularity, so the bigger the genre, the more popular it is on Spotify.
-> The word cloud shows that the most popular genres on Spotify are dance pop, electro house, hip hop, and electropop, as they are the largest and most prominent in the image.

Artist with most followers [7] :



-> This horizontal bar plot displays us the top artists with most followers in the 2020-2021 time period.
-> We have an option to select the number of artists that have to be considered for this plot using a slider.
-> The color also represents the total streams here the darker the orange the more streams.
-> For the above plot Ed Sheeran has the most followers in the time period with around 83,337,000 followers.

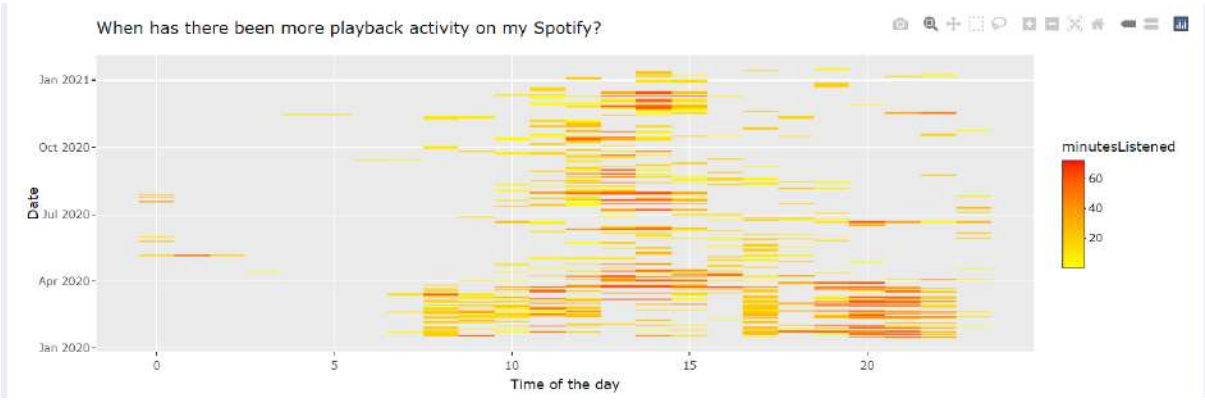
What weekday and time I have listened to most music [6] :



-> The heatmap chart is a visual representation of the amount of time the user has listened to music on Spotify across all 24 hours of every day.
-> The x-axis is the time of the day in 24-hour format and the y-axis is the weekday from Sunday to Saturday.

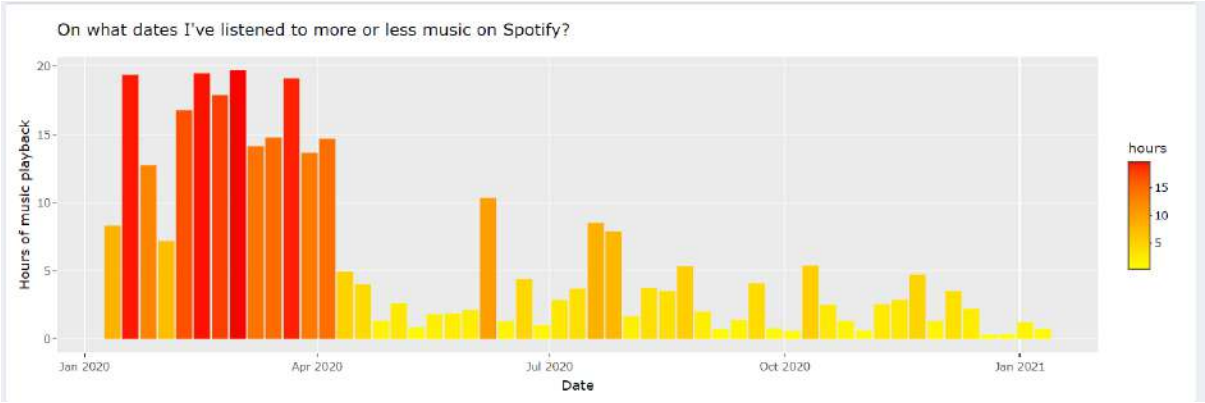
-> The chart is color-coded with yellow representing the least amount of time and red representing the most amount of time.
-> The chart shows that the user has listened to the most music on Friday and Saturday evenings, between 18:00 and 23:00, with more than 400 minutes each.
-> The chart also shows that the user has listened to the least music on Sunday and Monday mornings, between 6:00 and 11:00, with less than 100 minutes each.

Playback activity on Spotify:



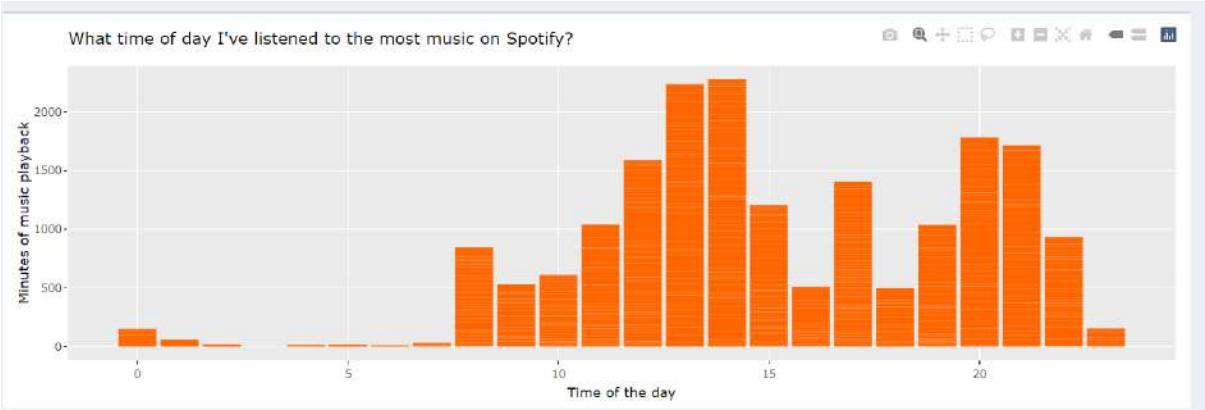
-> The x-axis is the time of the day in 24-hour format and the y-axis is the date from January 2020 to January 2021.
-> The chart is color-coded with yellow representing less minutes listened and red representing more minutes listened.
-> The chart shows that the user has more playback activity on their Spotify account during the evening and night hours, from 18:00 to 6:00, than during the day hours, from 6:00 to 18:00.
-> The chart also shows that the user has some variations in their playback activity throughout the year, such as more activity in February, April, July, and December, and less activity in January, March, May, and November.

Dates i have listened to spotify:



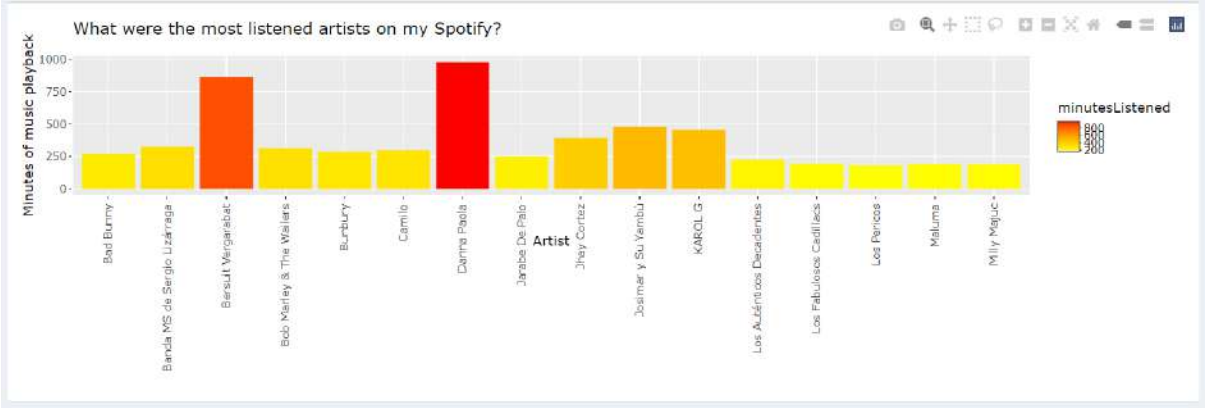
-> This is a vertical bar chart which shows on what dates the user has listened to spotify for more time by taking the number of hours listened into consideration.
-> Two hue colormap is used here with red and yellow.
-> This chart shows that the user has heard to the maximum amount of music in january 2020.
-> The least amount of music has been heard in the month of January 2021.

My most listened time of the day :



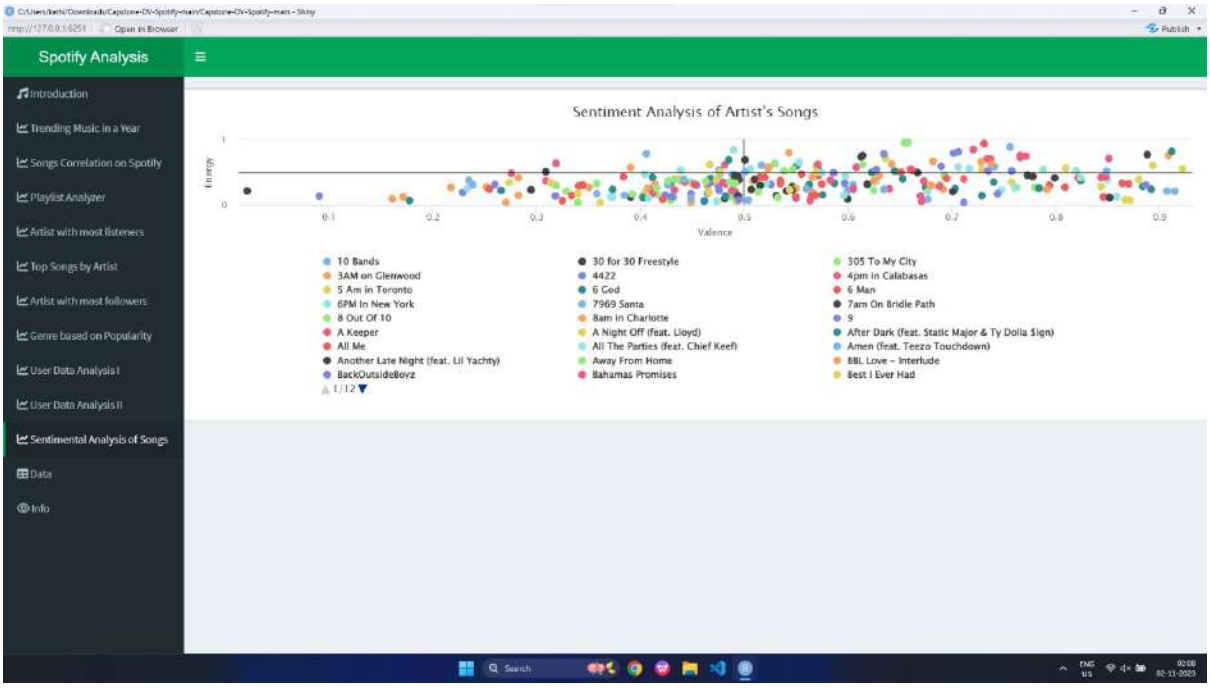
-> The stacked bar graph shows the amount of time the user has listened to music on Spotify by time of day and the songs listened are what is being stacked.
-> The x-axis is the time of the day in 24-hour format and the y-axis is the number of minutes of music played from 0 to 200.
-> The highest bar is at 20:00, with approximately 200 minutes of music played. This suggests that the user listens to the most music in the evening, possibly because they are relaxing, studying, or socializing.
-> The lowest bar is at 6:00, with approximately 0 minutes of music played. This suggests that the user listens to the least music in the early morning, possibly because they are sleeping, working, or busy.

My most listened artists on Spotify [9] :



->The bar graph shows the most listened artists on Spotify by the user.
->The x-axis is the artist name and the y-axis is the number of minutes listened from 0 to 600.
->The bars are colored red and yellow, but the colors do not have any significance.
->The highest bar is for "The Weeknd" with over 500 minutes listened. This suggests that the user is a fan of "The Weeknd" and enjoys his music style, which is a blend of R&B, pop, and hip hop.

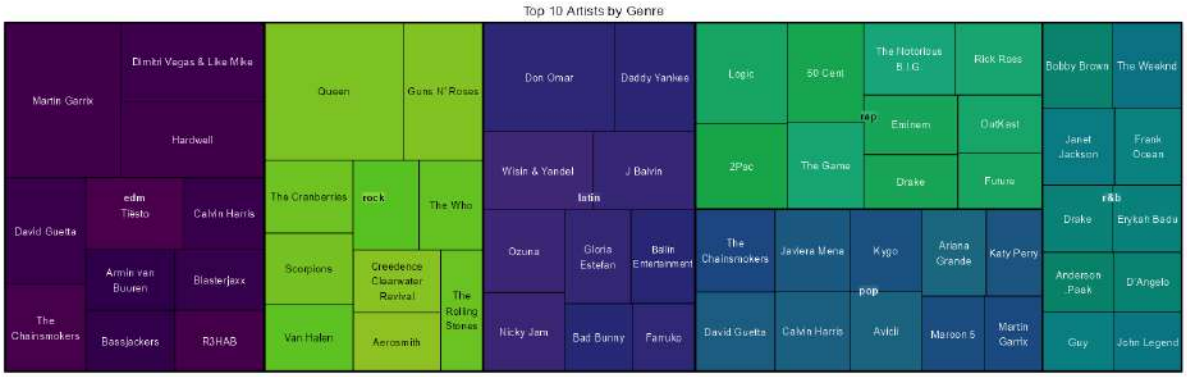
Sentimental Analysis [8] :



->The graph is a scatter plot that shows the sentiment analysis of artist's songs based on valence and energy.
->The x-axis is the valence, which is the musical positiveness conveyed by a song. A high valence means a more positive and happy song, while a low valence means a more negative and sad song.

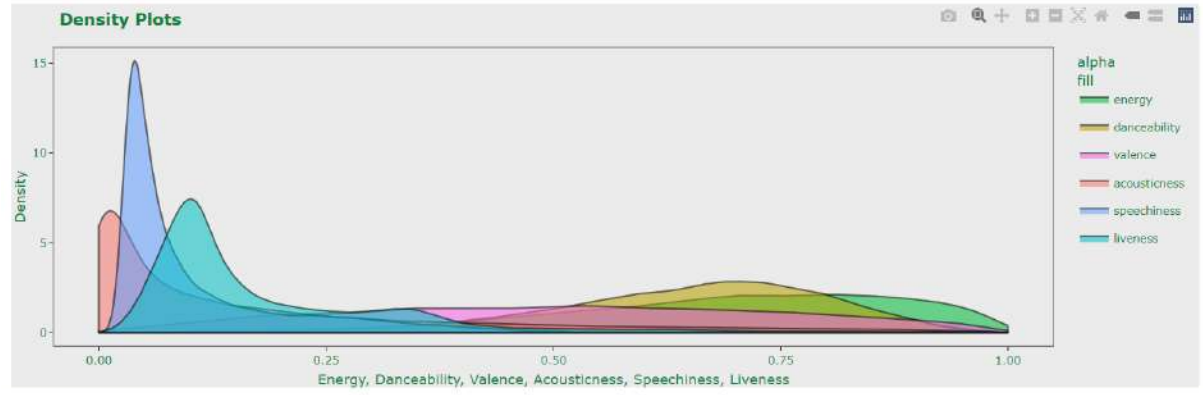
->The y-axis is the energy, which is the musical intensity and activity conveyed by a song. A high energy means a more energetic and fast song, while a low energy means a more calm and slow song.
->The graph has a large number of data points, each representing a song by the artist. The data points are colored according to the song name, which are listed in the legend at the bottom of the graph.

Top 10 Artists by Genre [14]:



-> The chart is a treemap that shows the percentage of music genres listened to by the user on Spotify.
-> The chart has 6 main divisions, each representing a different genre of music.
->The size of each genre represents how popular it is. And the size of each artist under it represents how popular the artist is.
-> The chart is color-coded by genre, with dance in blue, hip hop in green, rock in red, pop in yellow, R&B in purple, country in orange, jazz in brown, classical in pink, blues in gray, and folk in black.
->The chart shows that the user listens to a variety of music genres on Spotify, but the most popular ones are dance, hip hop, and rock, as they make up more than half of the chart.

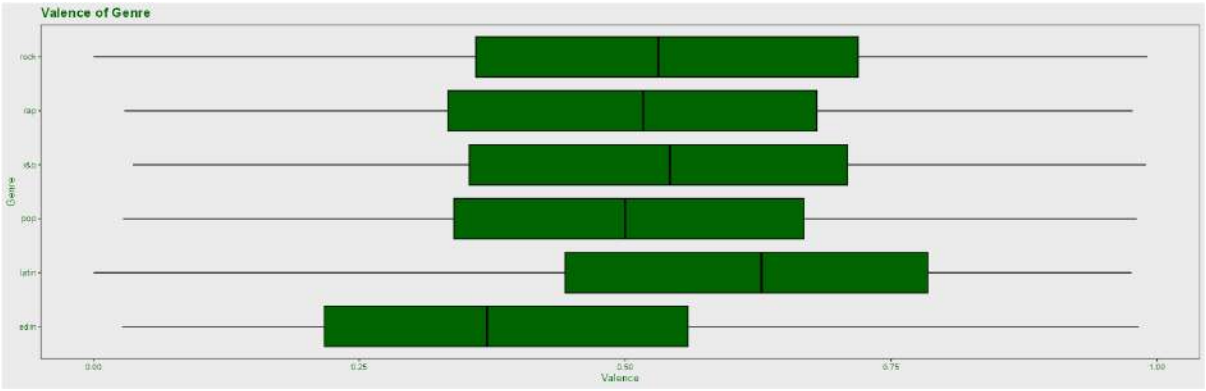
Density Plot [13] :



->The density plot shows the distribution of six musical attributes for different songs on Spotify.

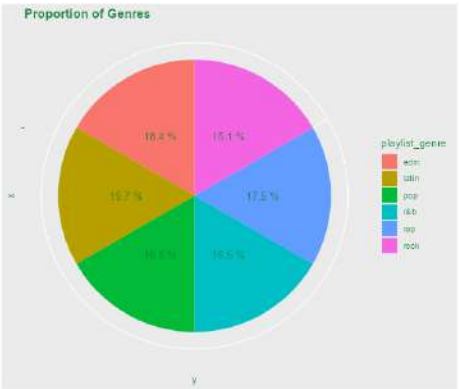
->The x-axis is the value of the musical attribute, ranging from 0 to 1. The y-axis is the density, which is the relative frequency of the value.
->The curves are colored by the song name, which are listed in the legend at the bottom of the plot. The curves are filled in with a lighter shade of their respective colors.
->The plot also shows that there is some overlap and similarity between the songs for some attributes, while there is some separation and difference between the songs for other attributes.

Box Plot for Valence of Genre [11] :



-> The box plot shows the distribution of valence for different music genres on Spotify.
-> The x-axis is the valence, which is the musical positiveness or happiness conveyed by a song. A high valence means a more positive and happy song, while a low valence means a more negative and sad song.
-. The y-axis is the genre, which is the musical category or style of a song. The graph has 6 genres: rock, pop, hip-hop, country, electronic, and metal.
-> The graph shows that rock has the highest valence among the genres, with a median of around 0.8 and a range of 0.4 to 1.2. This suggests that rock songs are generally positive and happy on Spotify, possibly because they are energetic, upbeat, and optimistic.

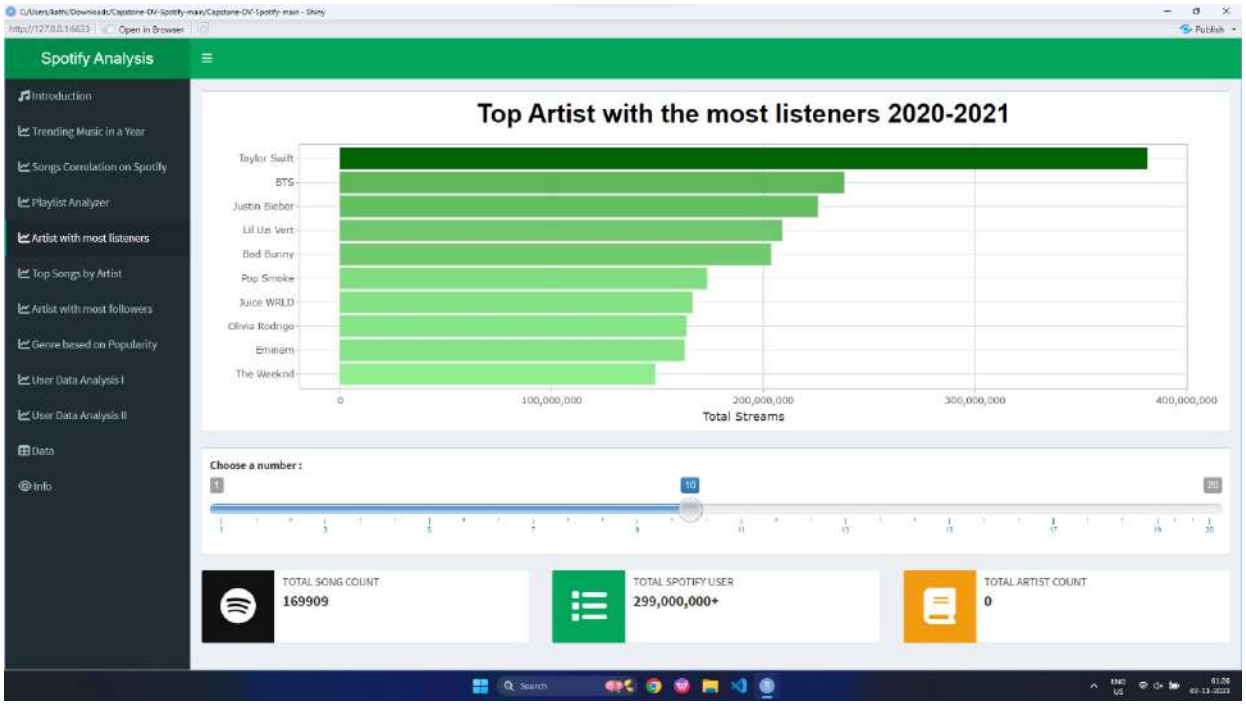
Pie Chart for Properties of Genre [2] :



-> The pie chart shows the proportion of music genres listened to by the user on Spotify.
The chart has 8 slices, each representing a different genre of music.
-> The chart is color-coded by genre, with pop/rock in blue, rap/hip-hop in green, country in red, R&B in yellow, electronic in purple, latin in orange, jazz in brown, and classical in pink.
-> The chart shows that the user listens to a variety of music genres on Spotify, but the most popular one is pop/rock, as it makes up more than half of the chart.
-> The chart also shows that the user listens to some less popular genres, such as latin, jazz, and classical, as they make up less than 10% of the chart.

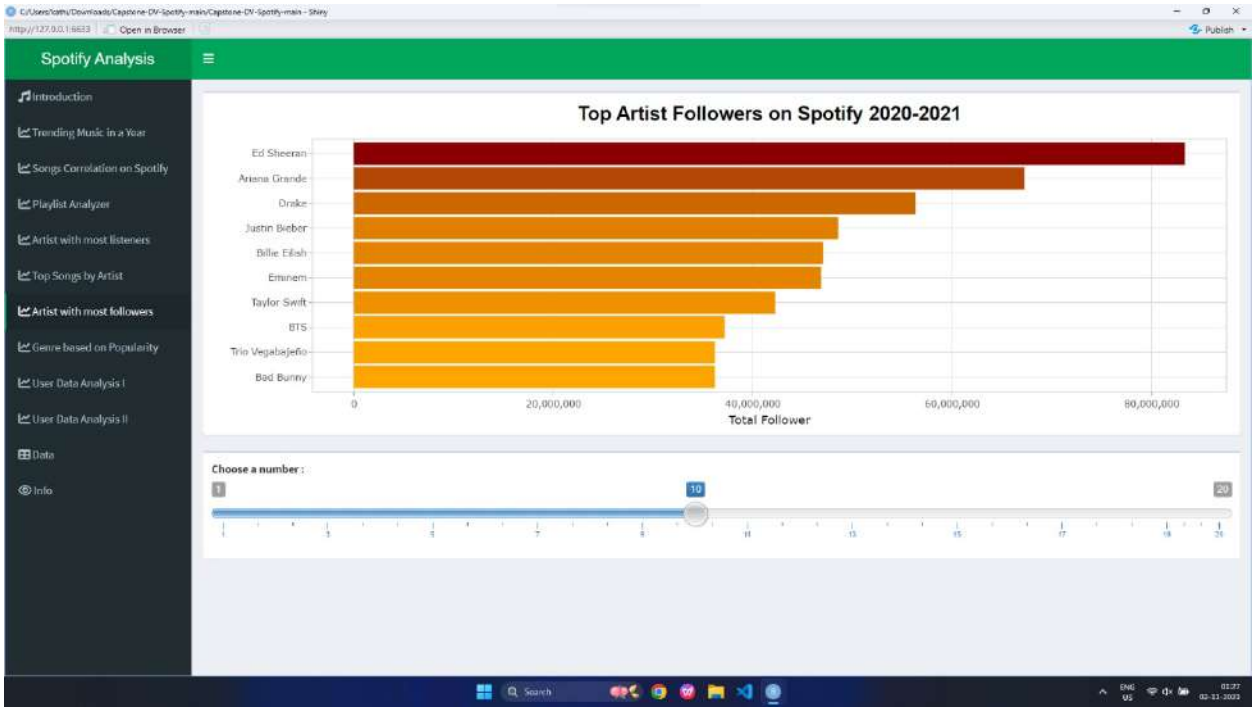
V. Analysis

Top Artist With Most Listeners



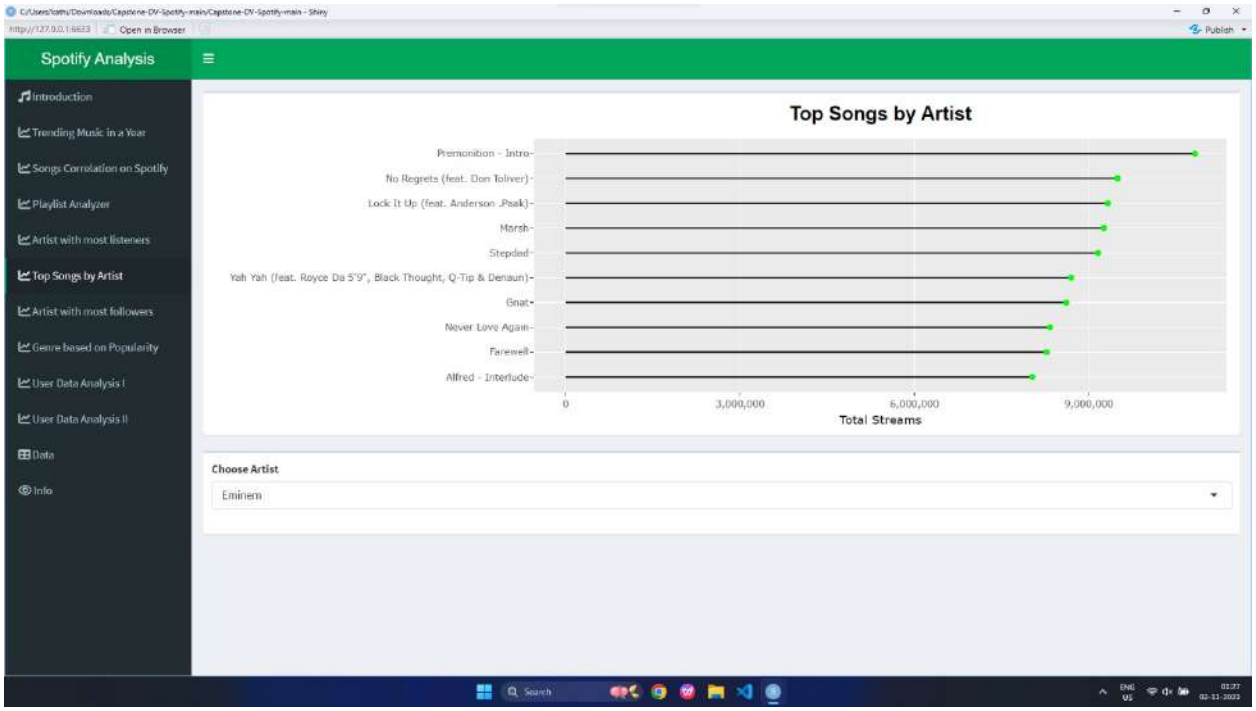
We chose to visualize the "Top Artists with the most listeners" using a horizontal bar chart because it provides a clear and concise way to compare the popularity of different artists based on their total streams. The horizontal layout allows for easy readability and comparison, especially when selecting a specific number of artists to display. The choice of color density from dark green to light green is a deliberate design decision. Green is often associated with growth and positivity, making it a fitting choice to represent the success and popularity of these artists. The transition from dark to light green not only helps users quickly identify the most popular artist at a glance, but it also visually communicates the relative differences in listener counts, reinforcing the hierarchy of success among the artists. This color scheme enhances the chart's interpretability and engagement.

Top Artist Followers



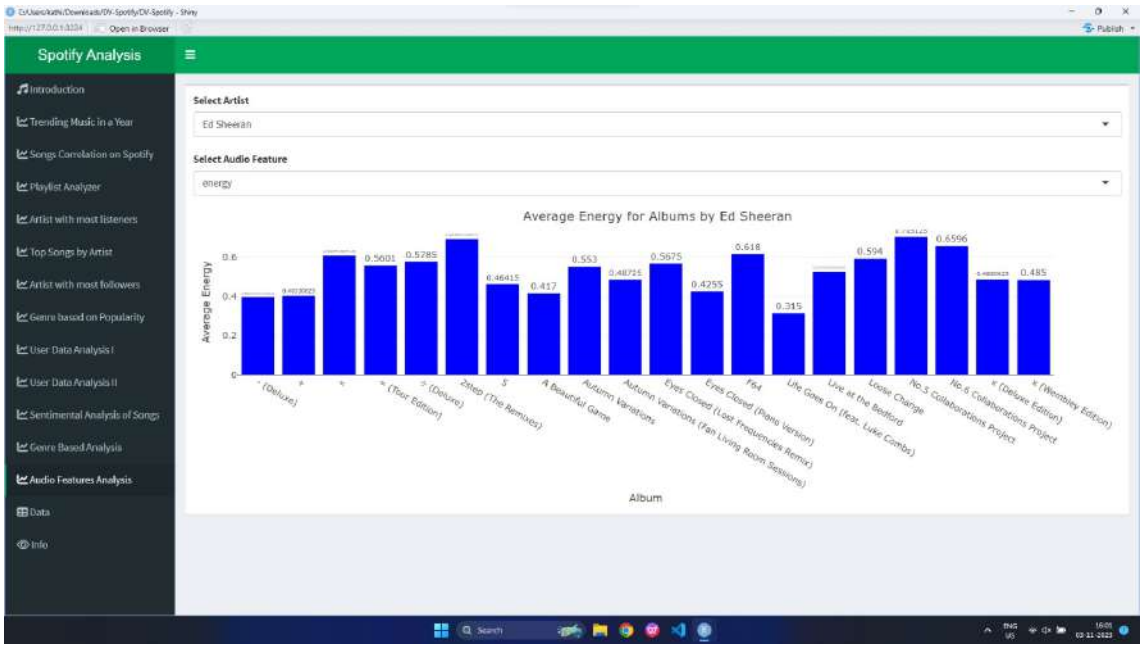
We selected a horizontal bar chart to visualize "Top Artists by Followers on Spotify" because it offers a straightforward means to compare artists based on their total followers. The horizontal layout is reader-friendly and accommodates flexibility in the number of artists displayed, ensuring the visualization's adaptability. The choice of color density from red to dark yellow to light yellow was intentional. Red often signifies prominence, making it suitable for representing high follower counts. Transitioning from dark yellow to light yellow serves to emphasize the differences in follower numbers, enabling quick identification of the most popular artists. This color scheme enhances the chart's visual impact, making it an engaging and informative tool for users to discern artist popularity at a glance.

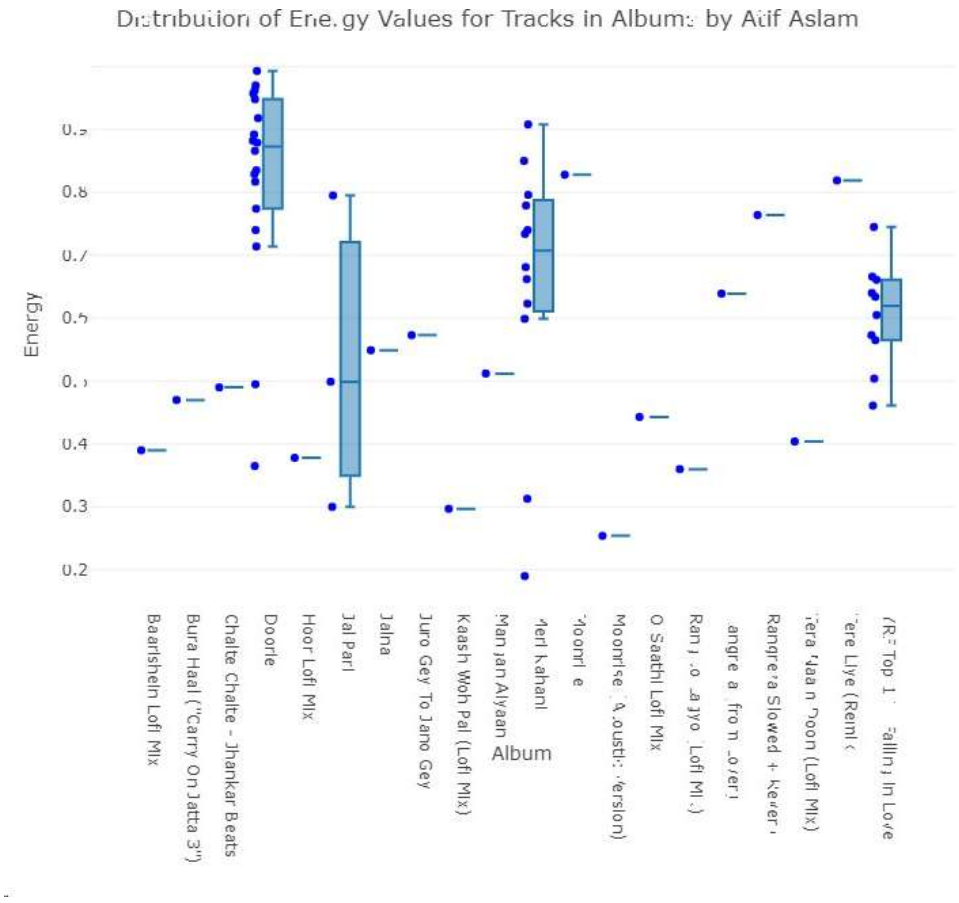
Top Songs by artist



We opted for a line chart to visualize "Top Songs by Artists" based on total streams because it's a dynamic way to track the performance of an artist's songs over time. This allows for easy selection of specific artists and visualizes their song streams chronologically. The addition of a point at the end of each line, indicating the total streams, provides a clear endpoint for understanding the cumulative popularity of an artist's top songs. This not only helps users spot trends in an artist's song performance but also offers precise and immediate insight into their overall streaming success in a user-friendly manner.

Audio Feature Analysis



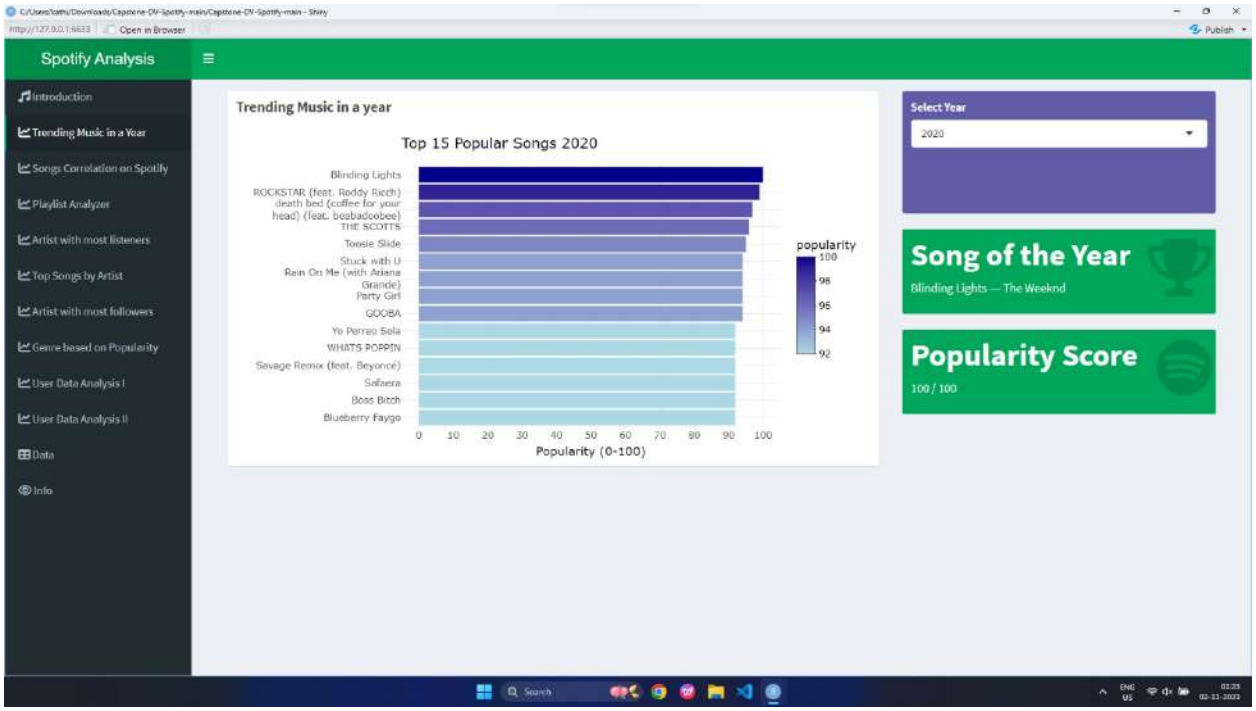


In our Spotify analysis project, we implemented a user-friendly vertical bar chart for audio feature analysis. This interactive visualization allows users to select an artist and an audio feature such as energy or acoustic ness from drop-down menus. The chart displays the average value of the chosen feature for each album by the selected artist. This analysis provides valuable insights into the artist's music style evolution across albums, aiding in understanding their creative trajectory and how it aligns with audience preferences. It also enables data-driven comparisons between artists, shedding light on their musical diversity and consistency.

A line chart is more advantageous than a box plot for visualizing the trend of music playback hours over time on Spotify. Line charts excel in representing trends and patterns by connecting data points with a continuous line. In the context of music playback hours, where time is a crucial factor, a line chart allows for a smooth and intuitive depiction of how playback hours vary over different time intervals, such as days, weeks, or months. This continuity is essential for showcasing trends and identifying patterns in user behavior, making it easier to observe fluctuations or growth in music consumption over time. Unlike box plots, which primarily emphasize the distribution of data, line charts are specifically designed to highlight temporal changes, offering a more dynamic and insightful representation of the evolving trends in music playback on the Spotify platform. The continuous line in a line chart provides a cohesive

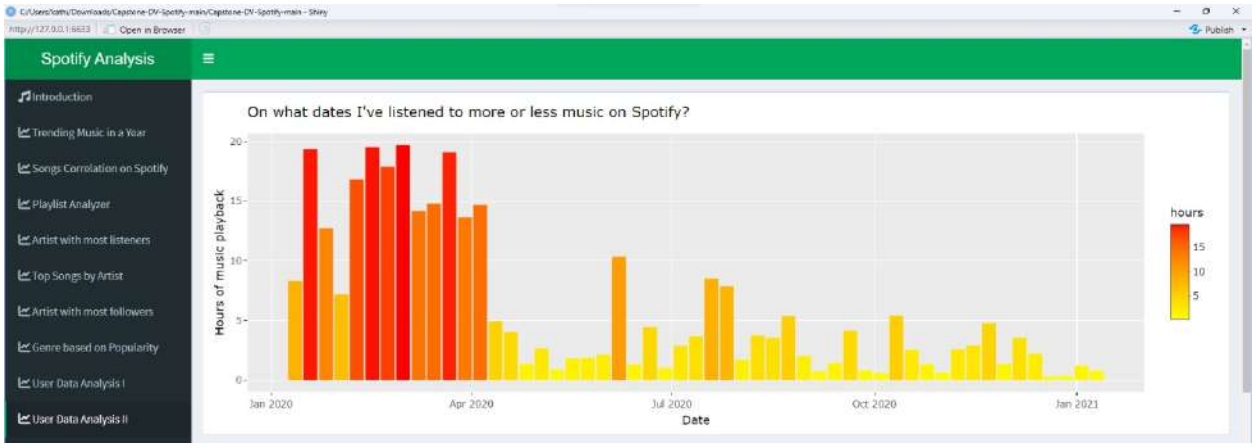
narrative, making it the preferred choice for visualizing the temporal aspect of music playback hours.

Trending music in a year



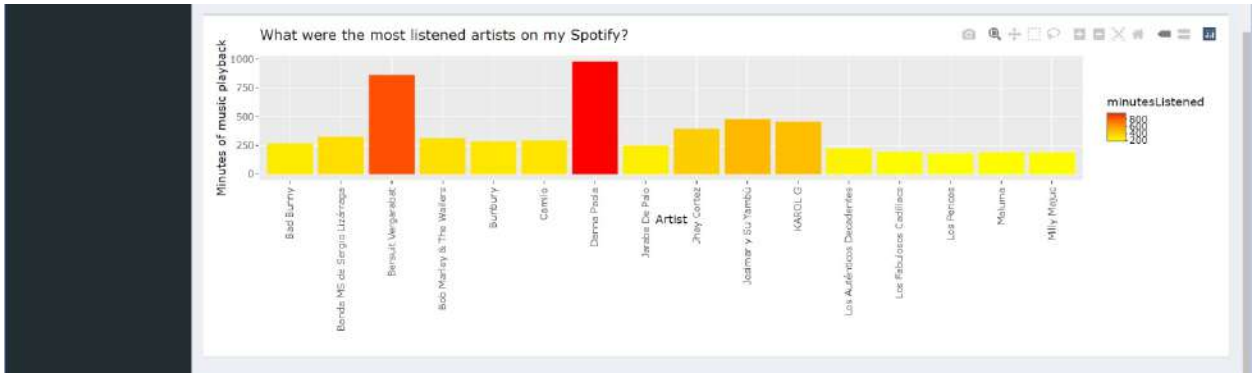
We selected a horizontal bar chart to visualize "Trending Music of the Year (Popularity Score)" because it offers a straightforward means to compare songs based on their yearly popularity. Users can conveniently select a specific year to see the trends in music over time. The choice of color density, transitioning from dark blue to light blue, was intentional. Dark blue, signifying higher popularity, is visually appealing and makes it easy to spot the most popular songs. The gradient from 100 to 92 emphasizes the variations in popularity scores. This visualization effectively showcases the "Song of the Year" and its corresponding "Popularity Score," allowing users to quickly identify the most trending songs for a given year while providing a clear representation of their relative popularity.

On what dates I have listened to more or less on Spotify?



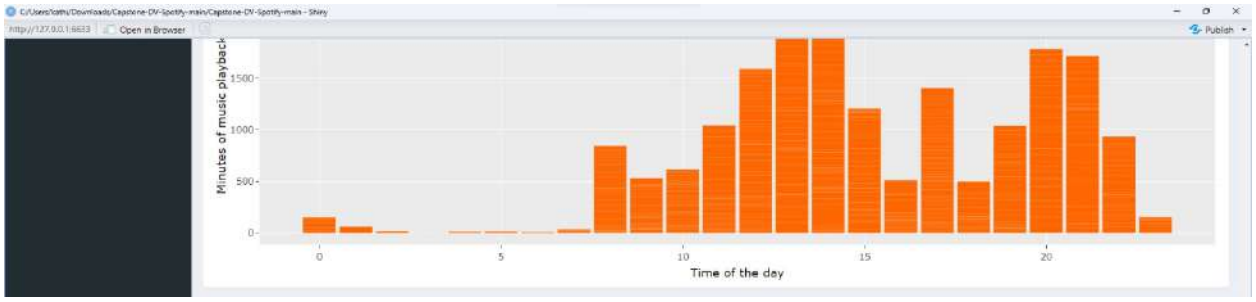
We decided to use a horizontal bar chart to show user's Spotify listening habits on various dates since it gives me an understandable and straightforward picture of how a user's music consumption changes over time. The horizontal arrangement makes it simple to compare dates and highlights trends and patterns in my listening habits. By adding a layer of information, color density as a metric allows one to easily determine which days have high or low listening activity. The data is easier to understand thanks to this visual method, which also helps us pinpoint the days when a user uses Spotify the most or least. This information can be useful for developing personal insights and optimizing user's usage going forward.

What were the most listened artists of my spotify?



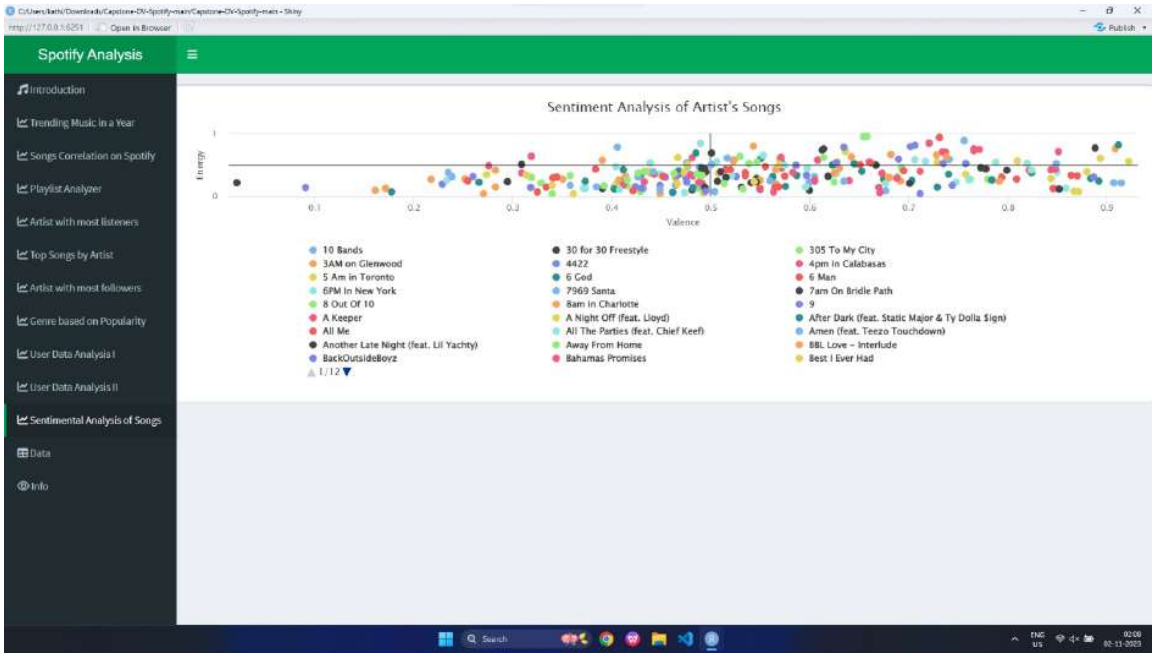
To provide you a clear and simple summary of your listening habits, we decided to use a vertical bar chart to show the artists that you have listened to the most on Spotify. It is simple to quickly recognize popular artists and trends with this approach. The transition from bright yellow to dark yellow in the color density is intended to imply development and importance. A graphic gradient that shows higher replay times for darker yellow and lower listening minutes for lighter yellow is a natural way to explain the hierarchy of artists' popularity in your music consumption.

What time of the day I've listened to the most music on Spotify?



A vertical bar chart based on minutes playback is how we decided to display the "What Time of the Day I Have Listened to Most Music on Spotify" data since it gives an accurate and succinct picture of a user's listening preferences. We can spot trends, determine the best times of day to listen to music, and use this visualization to inform data-driven choices that will enhance the user's musical experience. This data visualization makes it easy for us to know when the user is listening to music the most, which helps us make playlists and schedules that will make their Spotify usage more efficient and pleasurable.

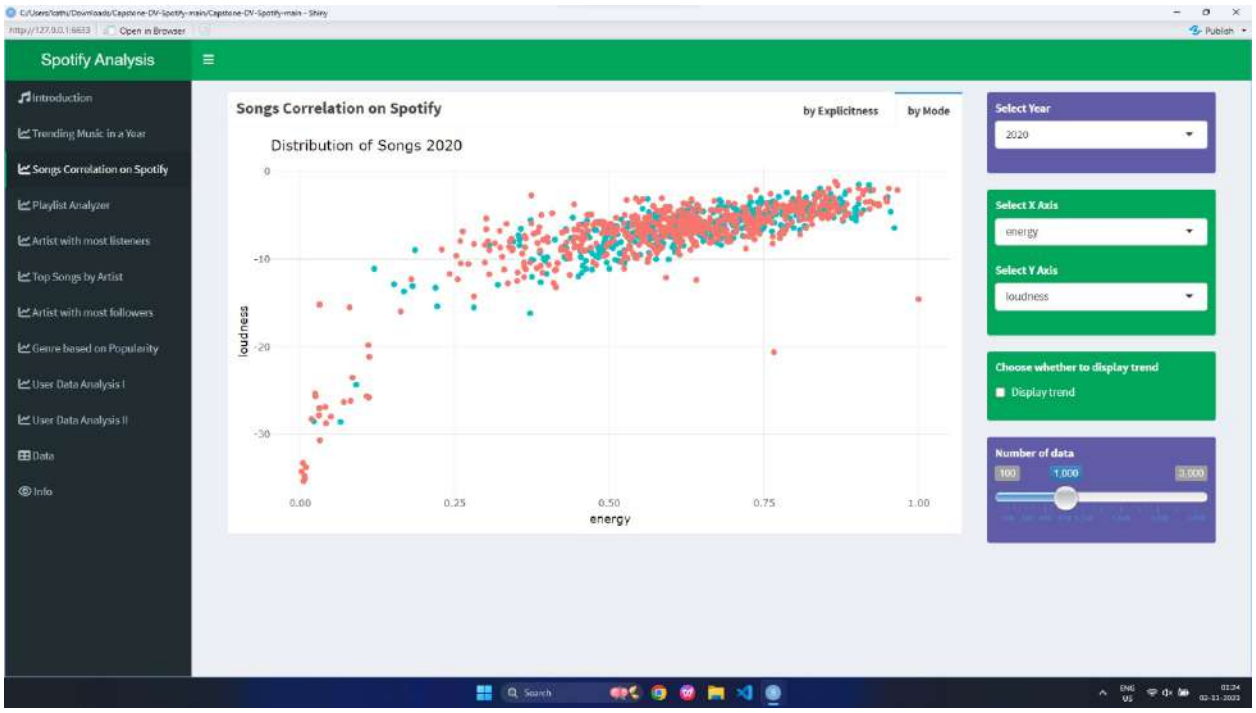
Sentimental emotion of the songs of the particular artist



We opted for a scatter plot to visualize "Sentimental Analysis of Artist's Song" based on energy and valence because it offers a comprehensive view of an artist's music, showcasing the emotional spectrum of their songs. Energy and valence are essential components of song

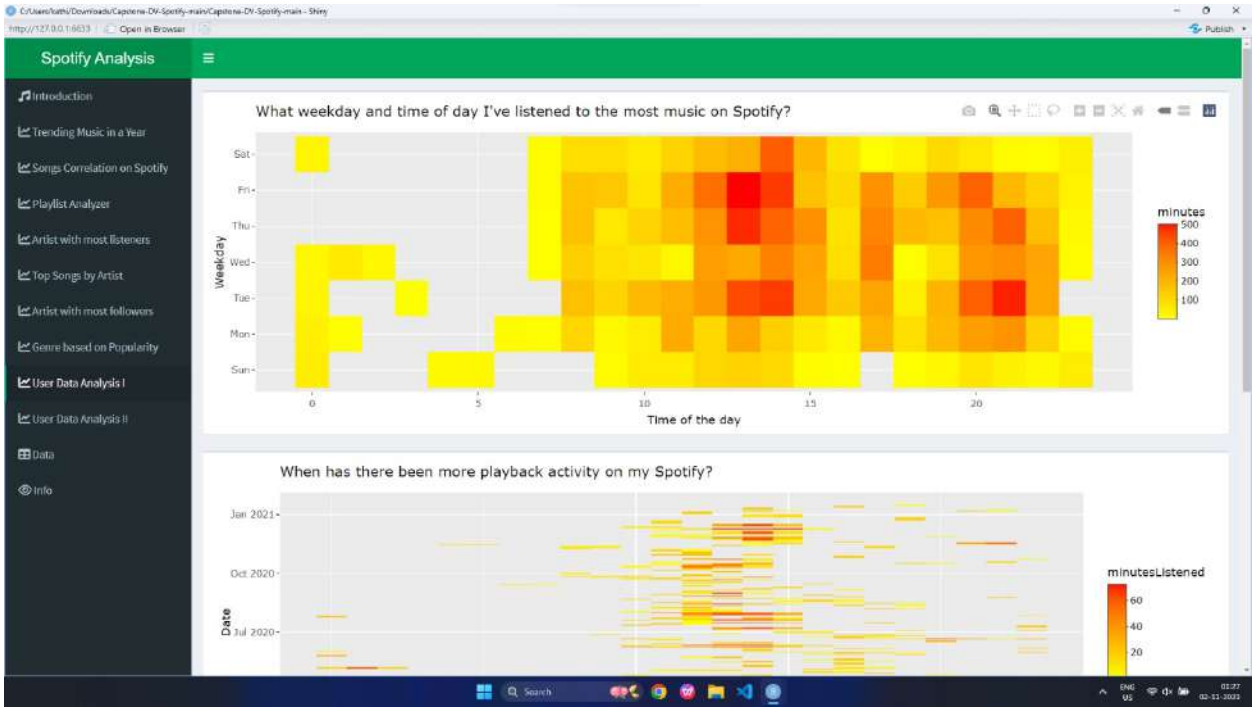
sentiment, and a scatter plot allows for the clear depiction of this relationship. Using different colors for different songs in the scatter plot is a deliberate choice. It helps users distinguish between individual songs and their unique emotional characteristics. Each color represents a specific song, allowing viewers to quickly identify and compare the emotional nuances across an artist's discography, making the analysis more engaging and informative for the audience.

Songs correlation on Spotify



We chose a scatter plot for "Songs Correlation on Spotify" using energy and loudness because it allows for dynamic exploration of the relationship between these musical attributes. Users can select the year, customize the axes, and choose data points to visualize, offering versatility in analysis. The use of two distinct colors, red and blue, is deliberate. It helps differentiate songs based on a specific characteristic, such as trend or popularity. Red signifies high values, while blue represents lower values, allowing users to discern correlations between energy and loudness at a glance. This color scheme enhances the chart's readability and provides a visual cue for understanding the strength and direction of the correlation between the selected attributes.

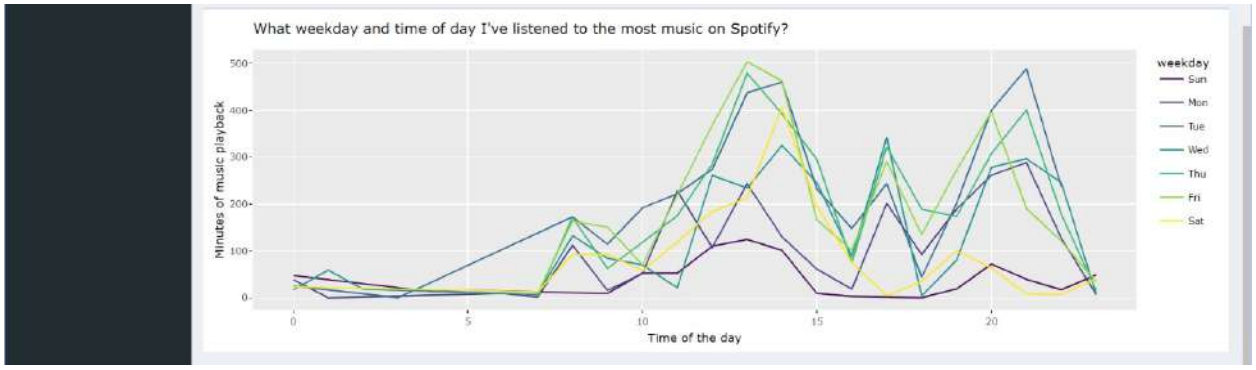
What weekday and time of the day I've listened to the most music on Spotify? & When has been more playback activity on my Spotify?



> The use of two distinct colors, red and yellow, is deliberate. It helps differentiate songs based on a specific characteristic, such as trend or popularity. Red signifies high values, while yellow represents lower values, allowing users to discern correlations between energy and loudness at a glance. This color scheme enhances the chart's readability and provides a visual cue for understanding the strength and direction of the correlation between the selected attributes.

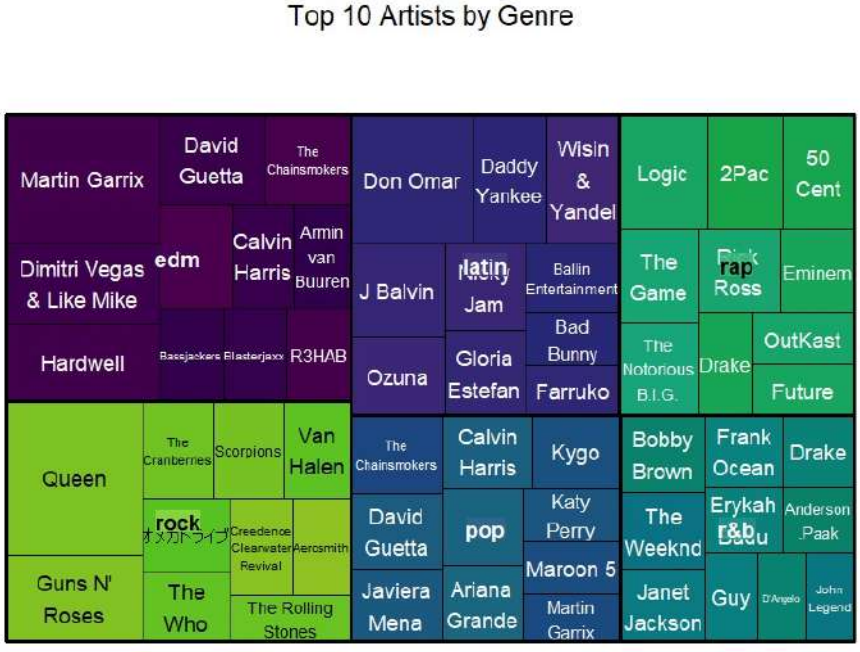
> This choice allows for a comprehensive display of playback intensity across various dates and times. The darker hues effectively indicate peak activity periods, while the lighter hues denote less active times. The transition from dark to light hues offers a clear visual representation of playback patterns. Additionally, the incorporation of labels or annotations adds a layer of precision to the visualization, enabling users to pinpoint specific dates and times with the highest activity. This visualization successfully fulfills the project's objective of providing an engaging and informative perspective on Spotify playback habits, making it a valuable tool for personal analysis and user engagement.

What weekday & time of the day I've listened to the most music on Spotify?



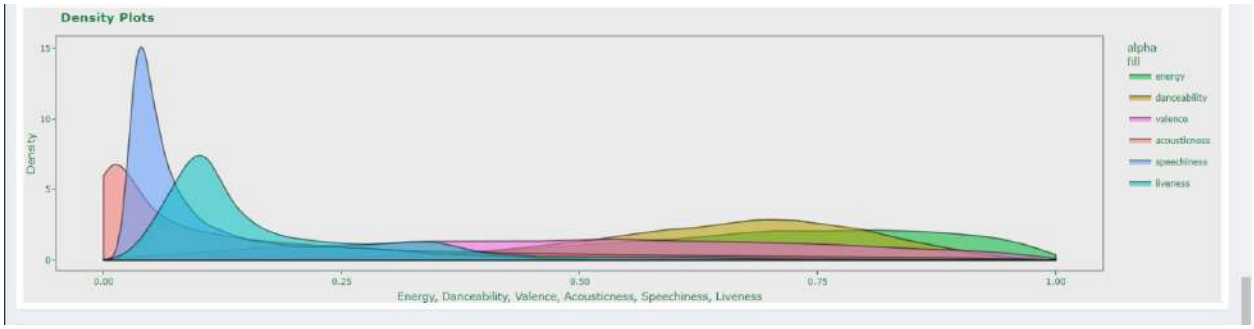
We find the decision to employ a line chart visualization to represent the user's Spotify listening habits based on weekday and time of day to be highly effective. This visualization allows for a clear, time-based depiction of my music listening patterns, with the x-axis representing time of day and the y-axis showing the minutes of music playback. The line chart provides a smooth, continuous view of my listening activity, making it easy to identify trends and daily variations. It effectively highlights the specific hours and weekdays during which I've listened to the most music. The choice of colors and line thickness can be thoughtfully selected to enhance the visual appeal and interpretability. Overall, this line chart offers a concise and engaging insight into my Spotify listening habits, allowing for a detailed analysis of the most active times.

Top 10 Artists by genre



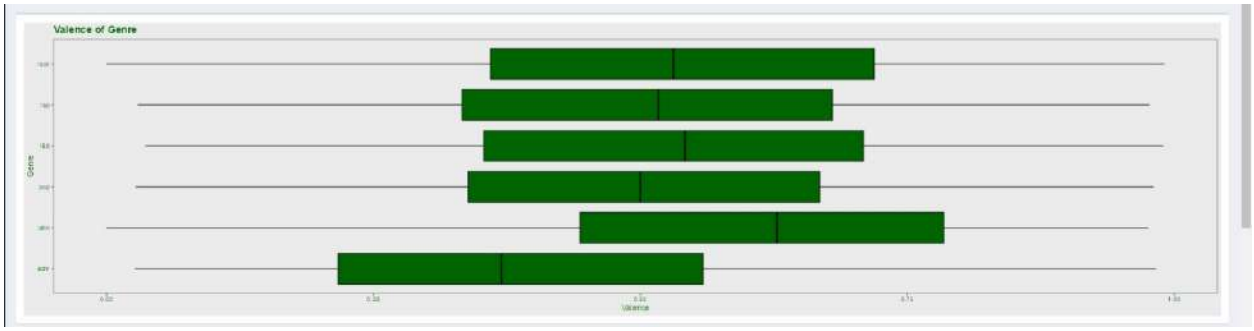
This visualization effectively displays the hierarchy and distribution of artists within different genres. Each genre is represented as a rectangle, with the size proportional to the number of artists it contains. Within each genre rectangle, smaller rectangles represent individual artists, and their sizes reflect their popularity within that genre. The tree map's visual hierarchy allows for a quick and easy comparison of genres and artists. This format offers an efficient and engaging way to showcase the diversity of artists across various music genres. Additionally, the use of color can be employed to further enhance the visual appeal and help users distinguish between genres. Overall, this tree map successfully conveys the top artists by genre in a visually engaging and informative manner, facilitating a comprehensive understanding of the music landscape.

Density plot of audio features



The density plot effectively illustrates the distribution and patterns of audio features, such as tempo, loudness, and instrumentalness, in a comprehensive and engaging manner. The density plot provides a visual representation of how these audio features are distributed, helping users understand the data's central tendencies and variations. The smooth curves and color gradients can be used to depict the density of feature values, allowing for a clear overview of the dataset's characteristics. By choosing this visualization, users can quickly identify the most common ranges for each audio feature and recognize any potential outliers or clusters. The density plot format ensures that the data is visually appealing, and it facilitates in-depth exploration and analysis of audio features. Overall, it enhances the project's capacity for users to grasp and interpret audio feature data effectively.

Box plot for genre based artists

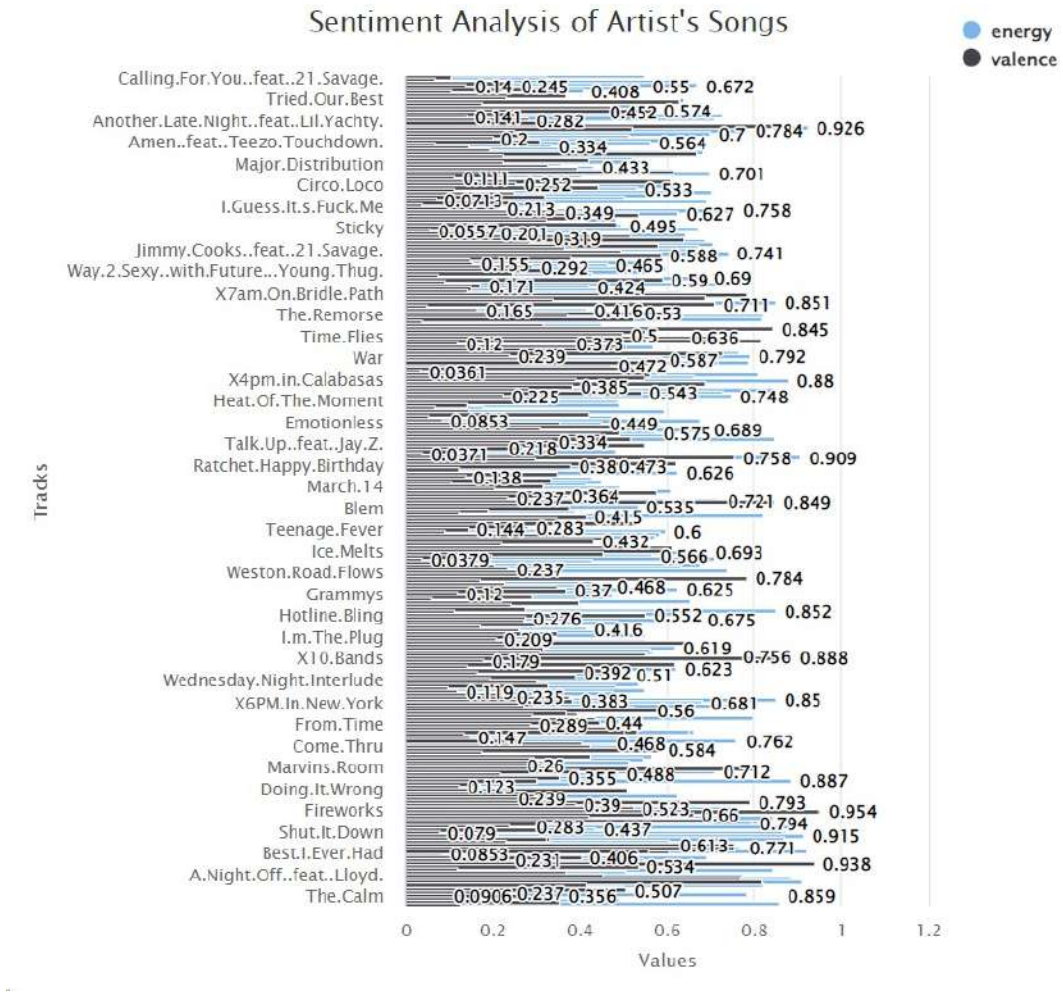


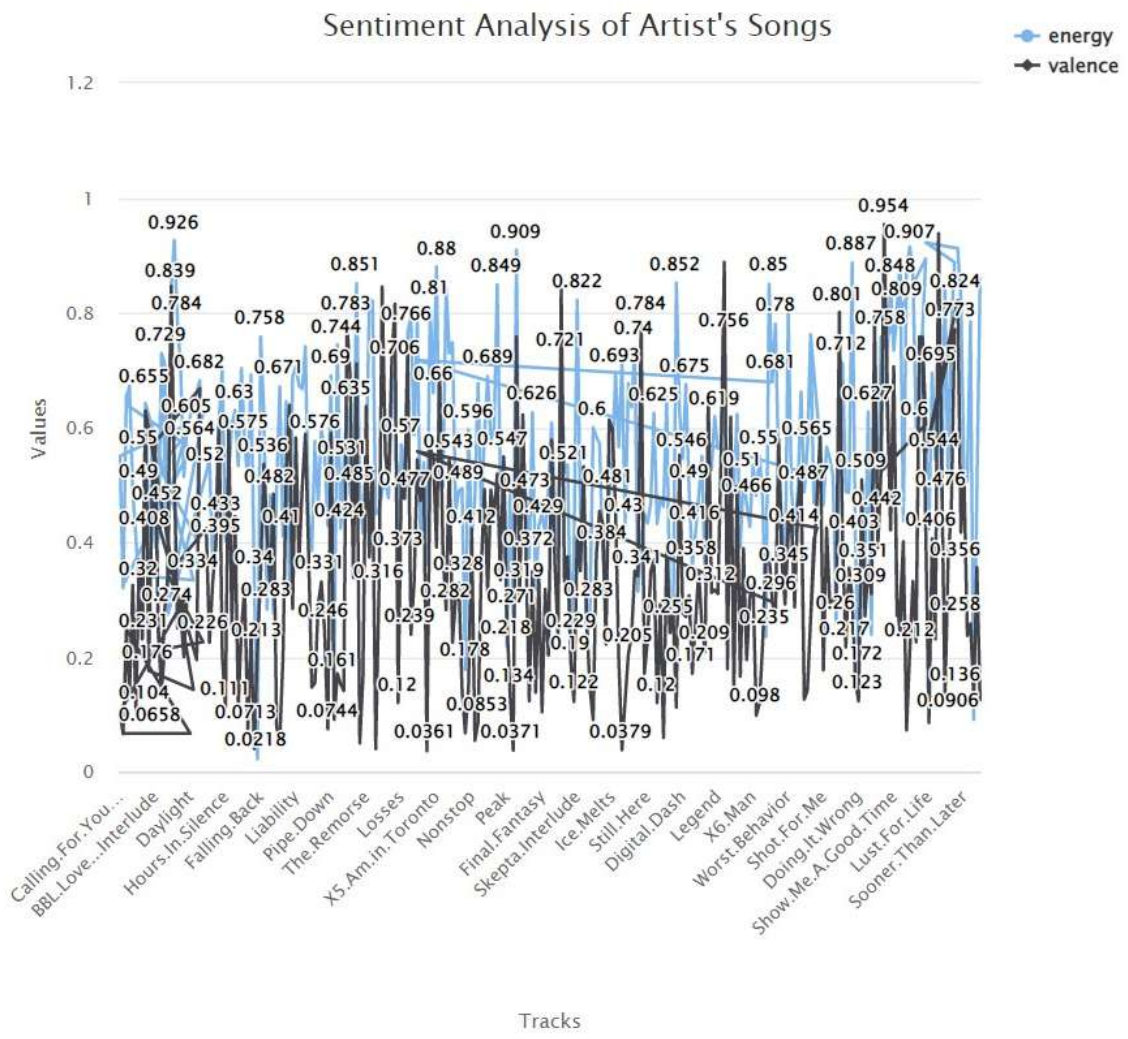
Box plots provide a clear and concise representation of the distribution of artist-related data within different music genres. The box plot design offers a visual summary of key statistics like median, quartiles, and potential outliers for each genre.

This format is particularly useful for comparing the central tendencies and variability of artist-related data across various genres. Users can easily identify differences in popularity or characteristics of artists within and between genres.

The project's design, which includes multiple box plots for various genres, allows for comprehensive genre-based analysis. Additionally, the use of colors and labels can enhance the interpretability of the data. Overall, the box plot visualization effectively conveys insights into the distribution of artists within different music genres, enabling users to make genre-specific comparisons and draw meaningful conclusions.

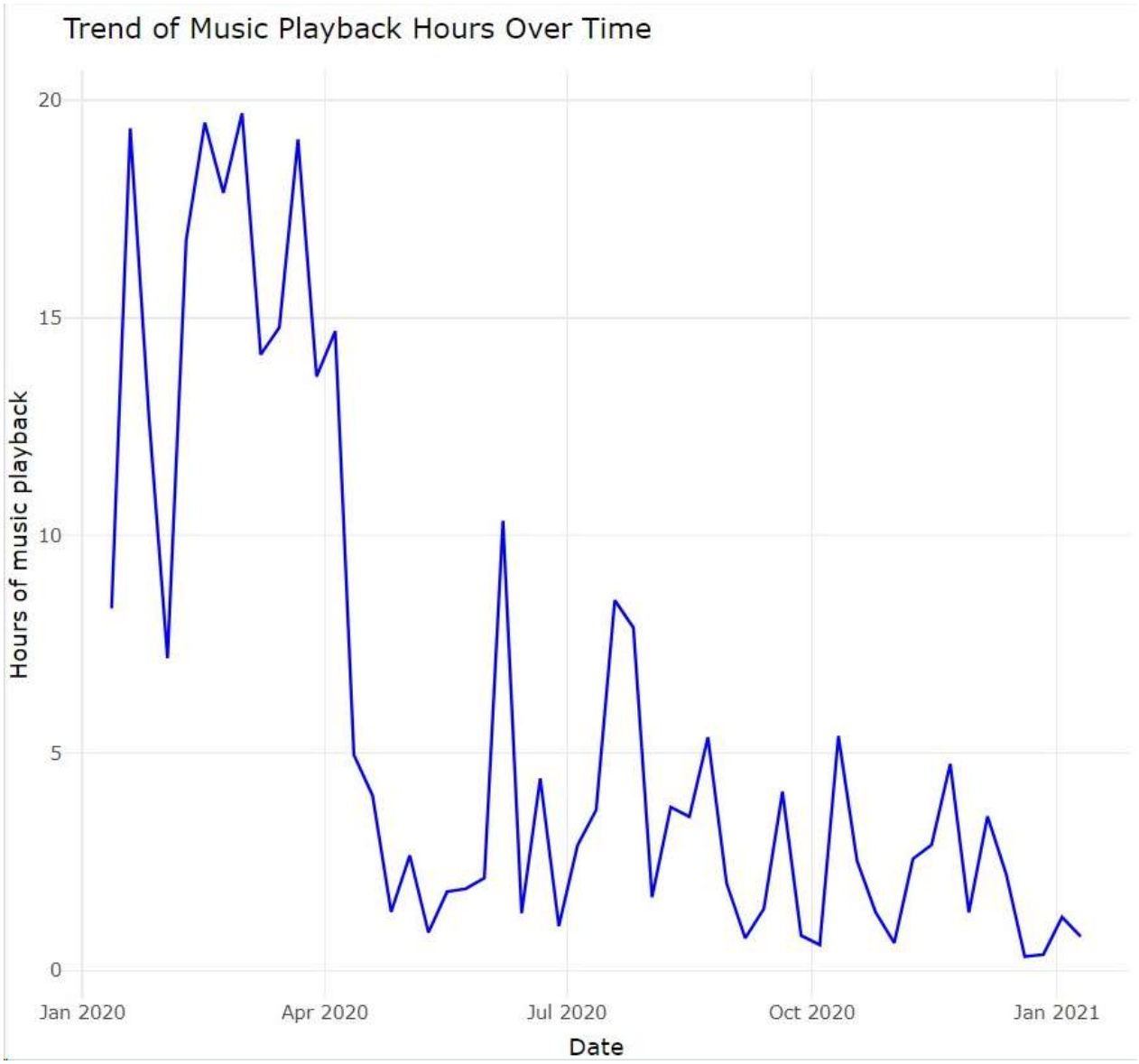
Sentiment Analysis

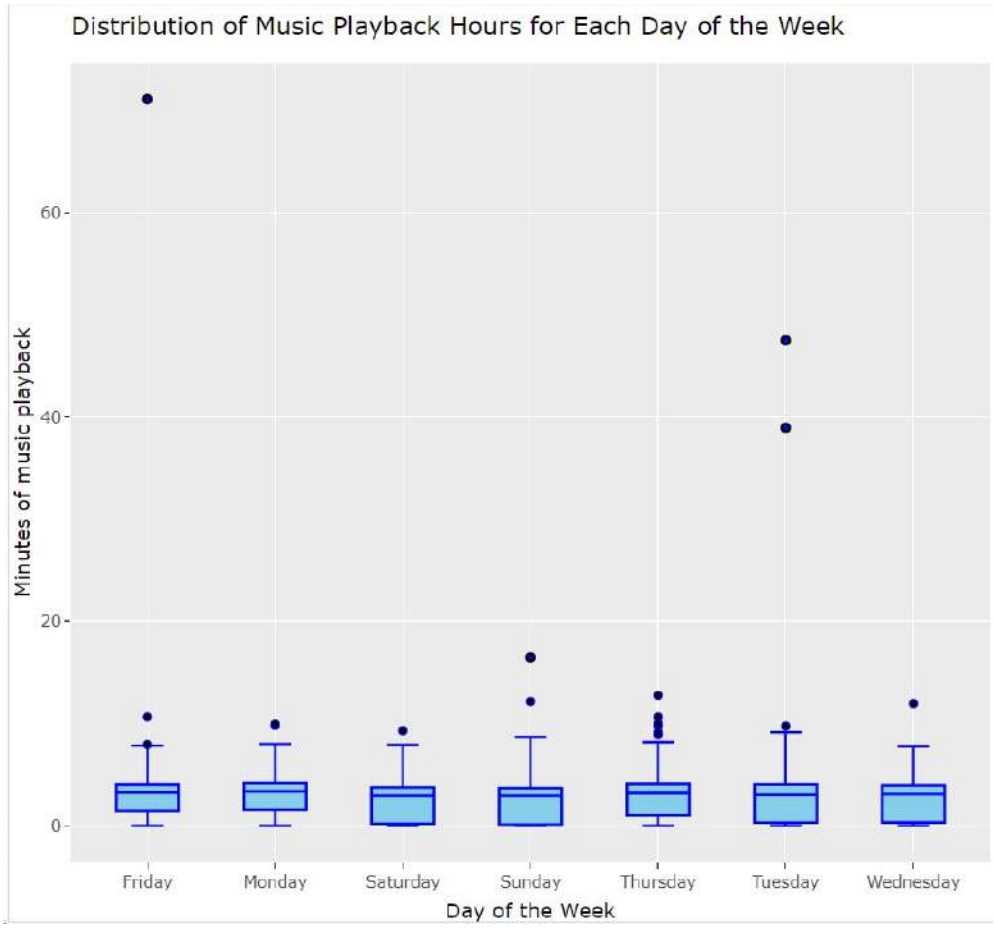




A bar chart proves advantageous for sentiment analysis of artists' songs on Spotify due to its capacity to effectively represent categorical sentiment data. Given that sentiment analysis often involves discrete categories such as positive, neutral, and negative, the distinct bars in a bar chart allow for straightforward comparisons across different songs or artists. The clear visualization of individual sentiment categories in separate bars enhances interpretability and accessibility, making it an ideal choice for conveying sentiment distribution and frequency. Bar charts excel in highlighting the prevalence of specific sentiments within a set of songs, providing a simple yet powerful means of communicating insights to a diverse audience. Their ease of interpretation, emphasis on discrete data, and suitability for categorical comparison make bar charts a valuable tool for visualizing sentiment analysis results in the context of music on platforms like Spotify.

Trend of Music playback hours over time





A line chart is more advantageous than a box plot for visualizing the trend of music playback hours over time on Spotify. Line charts excel in representing trends and patterns by connecting data points with a continuous line. In the context of music playback hours, where time is a crucial factor, a line chart allows for a smooth and intuitive depiction of how playback hours vary over different time intervals, such as days, weeks, or months. This continuity is essential for showcasing trends and identifying patterns in user behavior, making it easier to observe fluctuations or growth in music consumption over time. Unlike box plots, which primarily emphasize the distribution of data, line charts are specifically designed to highlight temporal changes, offering a more dynamic and insightful representation of the evolving trends in music playback on the Spotify platform. The continuous line in a line chart provides a cohesive narrative, making it the preferred choice for visualizing the temporal aspect of music playback hours.

VI. Conclusion and Future scope

In conclusion, the project "Exploring Musical Diversity on Spotify" has delivered a comprehensive and insightful examination of the musical ecosystem on the Spotify platform. It has illuminated the most influential artists with vast listener bases, offered a nuanced understanding of audio features that define genres and artists, delved into personal listening habits, and assessed the emotional sentiment of songs by specific artists. This project has not only celebrated the richness and diversity of music on Spotify but has also furnished valuable insights for music enthusiasts, industry professionals, and researchers.

It underscores the profound impact of the digital age on the music industry and the opportunities it presents for personalization and discovery. By unraveling the intricate web of music data and using interactive visualizations, the project has provided a dynamic and engaging journey through the world of music. Ultimately, it contributes to a deeper appreciation of the ever-evolving landscape of music in the digital era, and its findings will continue to inspire curiosity and exploration in the realm of music streaming.

Spotify visualization dashboard showcases insightful data trends, providing a comprehensive view of user preferences and music consumption patterns. Looking ahead, expanding the dashboard's capabilities to incorporate real-time analytics, personalized recommendations, and integration with social platforms could elevate user engagement. Incorporating machine learning algorithms for predictive analysis and leveraging AI to enhance music discovery might further enhance the platform's appeal. Collaborating with artists and integrating live concert updates could also enhance the user experience, ensuring sustained relevance in an evolving music landscape.

References

[1] Jingiting Li, Analysis of The Trend of Spotify, "*BCP Business & Management,* vol 34, pp. 919 - 926, Dec, 2022.

[2] Zaid Al Betawi *et al*, What Makes a Song Trend?, "*Journal of Marketing Development and competitiveness*, vol 14, pp. 78 - 91, 2020.

[3] Thanawat Chistri, Strategic marketing analysis : a case study of Spotify, " *Chulalongkorn University Theses and Dissertations*, pp 5-23, 2021,

[4] Jayachithra Kumar *et al*, Using Visualizations to Encourage Blind-Spot Exploration, "*IntRS Workshop, Vancouver, Canada*, vol 2225, Oct 2018.

[5] Bello, P., *et al*. Cultural Divergence in popular music: the increasing diversity of music consumption on Spotify across countries. *Humanit Soc Sci Commun 8*, 182, July 2021.

[6] Ivanova, Inna, *et al*, Explorify: A Personalized Interactive Visualization Tool for Spotify Listening History.

[7] Kamehkhosh, Iman, *et al* , Effects of recommendations on the playlist creation behavior of users." *User Modeling and User-Adapted Interaction*, vol 30, pp - 285-322, 2020.

[8] Liang, Yu, *et a*l. "Promoting music exploration through personalized nudging in a genre exploration recommender." *International Journal of Human–Computer Interaction ,* vol 39.7, pp 1495-1518, 2023.

[9] Aguiar, Luis, *et al*, Platforms, promotion, and product discovery: Evidence from Spotify playlists, *National Bureau of Economic Research*, no. w24103, 2018.

[10] Scarratt, Rebecca Jane, *et al*, The audio features of sleep music: Universal and subgroup characteristics, *PloS One*, vol 18.1, 2023.

[11] Fredsson, Dennis. "SpotiVis-Finding new ways of visualizing the spread of popular music.", 2021.

[12] Petitbon, Aimée M., David B. Hitchcock. "What Kind of Music Do You Like? A Statistical Analysis of Music Genre Popularity Over Time." *Journal of Data Science ,* vol 20.2, 2022.

[13] Duman, Deniz, *et a*l. "Music we move to: Spotify audio features and reasons for listening." *Plos one* , vol 17.9 , 2022.

[14] Hadash, Sophia, *et al*, How Playlist Evaluation Compares to Track Evaluations in Music Recommender Systems." *IntRS@ RecSys*. 2019.

[15] Guedes, Leandro *et al,* Exploring Music Rankings with Interactive Visualization *Proceedings of the Symposium on Applied Computing,* April 2017.