ANALYZING THE BILLBOARD TOP 100 IN THE 21ST CENTURY

Nathan Brito, Kaito Minami, Jeffrey Pan, Katie Warren

Northeastern University, Boston, MA, USA

Abstract

We have built a dashboard analyzing the top 100 Billboard songs in the 21st century. The dashboard allows the user to choose a range of years, display different visualizations for song features including artists, genres, and lyrics, and recommend other Billboard top 100 songs based on the user's input song. Data was gathered from a variety of sources, from the Billboard top 100 website to Genius, last.fm, and Spotify APIs. These sources were utilized to consolidate song information into a master dataset that powers the dashboard. From there, several dashboard elements were implemented to make the dashboard a fun and user-friendly place to examine trends in the music industry. The dashboard is available in the linked GitHub repository.

Introduction

The Billboard Top 100 is an iconic music chart that has been around since the 1950s. Over the years, it has served as a platform for measuring the popularity of songs and artists in the United States. In the 21st century, the Billboard Top 100 continues to be a valuable tool for understanding the trends and preferences of music consumers. From a nostalgia perspective, analyzing the Billboard Top 100 can bring back memories of the most popular songs and artists from a particular time period. Studying the Billboard Top 100 can reveal interesting patterns and insights into the music industry, such as which genres are popular, which artists are dominating the charts, and overall sentiment of songs over the years. The analysis of the Billboard Top 100 in the 21st century offers a unique perspective on the evolving landscape of popular music and its impact on culture.

To create a user-friendly platform where one can gain a better understanding of 21st century music, we created a dashboard using Plotly, a powerful Python library for visualizing data and presenting insights in an interactive and engaging way. By designing and implementing a dashboard to analyze Billboard's Top 100s throughout the 21st century, we can provide an immersive experience that allows users to explore the data and gain a deeper understanding of the trends and patterns that emerge. In addition, the dashboard allows for the inclusion of album covers and song samples, adding a unique perspective on the cultural significance of the top 100, and allowing users to experience the nostalgia and emotional sentiment of the music that has topped the charts over the years. We will be investigating song, artist, and genre trends for the latter decade of the 21st century, from 2012 to 2022. We believe that the most popular songs will yield higher sentiment scores, as happier songs usually lead to more public acceptance. We also believe that those top-chart songs stem from a range of genres and artists, and are not completely dominated by a genre. We will explore these working hypotheses with the use of our dashboard and analyze the visualizations in regard to this hypothesis.

Methods

A database was constructed by a combination of web-scraping methodology and manual entry. Since the Billboard Top 100 charts update on a weekly basis [1], a consistent list of historical songs was required to create a comprehensive analysis of the 21st century. A static list ensures that the data is accurate and reliable, and allows for meaningful comparisons to be made over time. With these songs, it is possible to identify trends and patterns that have emerged in the music industry, and to gain insights into

the factors that contribute to the success of different genres and artists. A blog website, “Billboard Top 100 Of,” provides a comprehensive list of top 100 songs for the entirety of the 21st century, allowing for us to obtain a consistent list of songs for each year [5].

Web-scraping methodology using the Beautiful Soup package was implemented on “Billboard Top 100 Of” to obtain artist name, song, and ranking for each year. A table was accessible within the HTML code of the website, and information for the majority of years was able to be extracted with the help of a for-loop changing the website URL. However, the layout of the website was not consistent across all years, as the blog had implemented new features such as song album covers. These new features contained normal text paragraphs of the song information and were not housed in a table, therefore making the scraping process for those years difficult. Years that did not have a table in their HTML structure were manually entered and validated by a member of our team.

Once the dataset was completely assembled with artist names, song titles, and rankings for each year, further information was extracted using the following APIs:

- Genius
- Spotify
- last.fm

Using song title and artist name in combination with the three APIs, song lyrics [2], song characteristics such as danceability, energy, and valence, and genres [4] were obtained and stored in separate columns in a Pandas dataframe. Danceability was later sorted into an extra categorical column where it was partitioned into 5 categories, from least danceable to most danceable. Song samples and album covers were notably provided with the Spotify API [6], and were later implemented into the dashboard to provide users with a more interactive experience. Genres were returned in the form of lists, with multiple different genres being listed per song. For easier analysis, a member of the group validated each genre and applied a song to a “general genre.” A situation of this can be the genre “rock” being assigned to a song that has a genre list that contains “pop rock” and “soul rock.”

The Billboard Top 100 contains a broad variety of songs, with varying profanity. To be as user-friendly as possible, profane words were removed from the lyrics using the “ProfanityFilter” Python package. This removal of words may influence the sentiment of certain songs, but consideration of users is prioritized in this project implementation.

With a fully constructed dataset, we began to visualize relationships between artists, genres, and lyrics. To provide users with insights in relations between artists and time, Sankey diagrams providing information on the danceability of artists and bar charts depicting artist appearances on the Billboard were constructed. To provide users with insights between genres and time, Sankey diagrams depicting relationships between artists, genres, danceability and years were implemented. Years serve as the middle layer for the Genre Sankey. A line plot showing different genre trends, a scatter plot visualizing the genres of the songs in a ranked order, and a horizontal bar chart with the top 10 artists that per genre were also created. To learn more about song lyrics, average sentiment over time and top song sentiments were visualized for the user to analyze.

Each of these visualizations have the ability to be filtered by a range of years, with a start year and an end year section in the dashboard. As the user customizes the year range, each visualization will be filtered by the start and end year values, and will provide the user with information on songs within those years. This feature allows users to analyze certain periods of time that they want to explore or hold dear to their hearts.

In addition to the visualizations, two user-interactive features were added to further satisfy their needs. The first feature is a menu that lists top k songs each year, where a slider allows the user to show the top k songs of the selected year range. With these songs, a picture of the album cover is provided alongside a sample of the song that can be played. Alongside this capability, a song recommendation system that implements the K Nearest Neighbors algorithm is also included, and recommends 5 songs based on a user-selected song. The features used in the k-nn algorithm are: sentiment score, danceability, energy, key, loudness, mode, speechness, acousticness, instrumentalness, liveliness, valence, tempo, duration in ms, and time signature. Genre was in consideration as a feature, but since some songs have 'unknown' primary genres and the number of genres for each song varies, we settled on these features and the calculated sentiment scores as the features for neighbor selection. The recommended songs also include the album cover and song samples, which allows the users to easily get to know more about the similarity and history of these songs.

Analysis

To demonstrate the power and capabilities of the Plotly dashboard, the visualizations mentioned above were created with a default year range of 2012 to 2022. The default year range of 2012 to 2022 was selected for its broad range of genres and artists. A broader range can provide more insights into artists, songs lyrics, and genre compared to a smaller one.

Within the "Artist Insights" tab on the dashboard, we created a visualization of the top 10 artists based on their popularity between 2012 and 2022 (Figure 1). This chart provides valuable insights into which artists were the most popular over the last decade. According to Figure 1, Drake has the highest number of songs on the Billboard Top 100 in the past decade, with 32 songs being on the list. The runner up, Maroon 5, has only 17 songs. The visualization provides insights on the resilience of artists, including Justin Bieber and Taylor Swift, who have 13 and 15 songs respectively and have produced hit songs continuously over the years. While these insights provided by the visualization show evidence that top-chart songs are mainly produced by popular artists, there are many one-hit wonders that make up the ranks of the Billboard Top 100. Some artists will receive more exposure than others, especially when they are more popular, which can explain Drake's whopping 32 songs in the past 10 years. Coupled with Figure 2, "Top Artists by Genre," 6 of the top 10 artists are pop singers, showing the genre's dominance throughout the past 10 years. With these singers, one can determine the danceability of the top artists with 12 or more hit songs within 2012 and 2022 by using the "Artist Sankey" feature (Figure 3), and can gain further insights as to whether or not popular artists produced more danceable music or not during their fame. According to the Sankey, artists that had more popular songs often had more or most danceable songs.

In the "Lyric Insights" tab on the dashboard, one can visualize the average sentiments of the top 100 song lyrics each year within the given time frame. From 2012 to 2022, each year has had a positive sentiment score, ranging from 0.06 to 0.11 (Figure 4). However, since 2013, there has been a decrease in overall sentiment to 2019, with 2017 being the exception. This decrease in sentiment can be attributed to the "downward trend in 'happiness' and 'brightness', as well as a slight upward trend in 'sadness'" [3]. Even with profanity removed, a downward trend in sentiment can be observed, providing valuable information on lyrical trends over the years. To further examine the sentiment of top songs, there is a sentiment color palette that represents the sentiment of the number one songs within the selected time period (Figure 5). This visualization helps contextualize the sentiment scores by assigning colors to the songs on a color scale where the songs with more positive lyrics are displayed in brighter colors as opposed to their more negative counterparts. When hovering, the user can see which years these songs are from. In this palette, it is easy to see that many of the number one songs are close to neutral sentiment, but there are a couple of more polarized number one songs. In 2021 and 2014, there were relatively positive number one songs such as "Happy" by Pharrell Williams and "Levitating" by Dua Lipa. Going outside of

the 10 year range, in 2006, the number one song had more negative lyrics in “Bad Day” by Daniel Powter (Figure 5).

Genre trends could be analyzed and determined with a combination of visualizations housed within the “Genre Insights” tab. Within the “Genre Sankey” figure, pop has by far been the most popular genre within the last decade for the top artists, followed by hip hop and rap (Figure 6). Genre trends are further amplified by observations found in “Genres Observed over Time,” which show interesting insights such as the country's gradual rise in popularity since 2016, hip hop’s decrease in popularity, and pop’s dominance in the music industry, despite its gradual decrease since 2017 (Figure 7). These trendlines can provide clues to what genres the general population enjoys, and which genres have the possibility of being even more popular in the next year. To dive deeper into the distributions of song genres across the Top 100 lists, the “Rank by Genre” visualization found within the same tab allows for users to visualize the rankings that each song has and its respective genre. In 2018, hip hop had a resurgent year, with the majority of songs between the top 20 to 50 being songs within that genre. However, in recent years, hip hop songs within the Top 100 are dwindling, with only a few making it into the top 20 songs and none in 2019 (Figures 7 and 8). By looking at Figures 7 and 8, we can see that despite pop being the most popular genre within the Billboard charts, there is still a diverse selection of genres within the top 100 ranks, supporting our hypothesis that the most popular songs are from a range of genres and artists.

One can further learn more about songs that one enjoys by employing the song recommendation system housed within the dashboard. If a user would request a song such as “Highest in the Room,” a rap song by Travis Scott, the recommendation system returns songs such as “Say It” by Tory Lanez or “Flashing Lights” by Kanye West. Both are known rap artists with songs with similar characteristics to “Highest in the Room,” and samples are available to be played by the user (Figure 9).

Conclusions

The dashboard constructed to analyze the Billboard Top 100 songs in the 21st century provides a plethora of insights, whether it be into artists, lyrics, or genres. A deep dive into a song’s past and its attributes allows users to feel nostalgic or discover new songs with the dashboard’s recommendation system and visualizations. The visualizations offered on the dashboard give users the ability to discover and learn more about popular artists, songs, and genres within a selected time period. Each visualization allows users to further specify queries by providing drop-down menus or interactable legends, enabling them to dig deeper into the characteristics of the songs of interest. For the scope of only the 21st century, the dashboard provides an intuitive approach to visualizing artist, song, and genre relationships.

With the analysis of Top 100 billboard songs between 2012 to 2022, we found that pop genre songs have dominated the decade, and learned that country music is slowly making a comeback from 2016. Drake has been a consistent top-chart artist, and referencing the trends and patterns generated by the visualizations, his success could be attributed to his continuous release of “danceable” music. In addition to “danceable” music being a potential driver for top-chart songs, the sentiment of the lyrics can influence a song’s performance with the public. Number one hit songs between 2012-2022 somewhat lean in the positive side, but historical songs such as “Bad Day” in 2006 can be very negative, and still be a number one hit song. This provides slight evidence to support our hypothesis that happier songs will be more popular. Consumers want to feel good about themselves and the music they listen to, and the majority of the top songs between 2012-2022 reflect this.

Users have the ability to listen to samples of songs recommended to them by the dashboard’s recommendation system. This capability allows them to explore the similarities and differences between the chosen song and the recommended ones, creating the opportunity for new songs to be discovered. While the recommendation system works in the scope of the songs of each year within the Billboard Top

100, for a more intuitive model with closer recommendations, a larger dataset containing songs outside of the list should be utilized.

While 22 years worth of top-chart music is only the tip of the iceberg in the music industry, the tools used to analyze it are powerful and can be applied across a much broader range of time, providing users and businesses with the insights and opportunities to discover new music and trends.

Author Contributions

The team was composed of four members, each with their respective roles and contributions to the project. Jeffrey was responsible for web scraping and manual data entry of the Top 100 Website, as well as writing up the findings and running analyses. He was also responsible for the cleaning of data lyrics, song titles, and artist names. Katie designed the dashboard, developed the song sampling and recommendation system, and created visualizations to be implemented within the dashboard. She was also involved with the analysis of visualizations. Kaito was in charge of collecting data through the API and creating visualizations with the information provided by them. He also contributed to the dashboard design and adapted visualizations to be implemented within it. Nathan collected and cleaned data through APIs and created visualizations while helping with the write-up. Together, the team worked on the project's poster presentation, showcasing their findings and insights to curious students at the poster fair. The team worked and communicated efficiently and effectively to provide a quality dashboard.

References

1. Billboard. (2022, July 5). *Frequently asked questions*. Billboard. Retrieved April 18, 2023, from https://www.billboard.com/frequently-asked-questions/
2. Genius. (n.d.). *Genius API documentation*. Genius API. Retrieved April 18, 2023, from https://docs.genius.com/
3. Iteriano, M., Kazemi, K., Wang, L., Yang, J., Yu, Z., & Komarova, N. L. (2018, May 16). *Musical trends and predictability of success in contemporary songs in ...* The Royal Society Publishing. Retrieved April 18, 2023, from https://royalsocietypublishing.org/doi/10.1098/rsos.171274
4. Last.fm. (n.d.). *Api docs*. Last.fm. Retrieved April 18, 2023, from https://www.last.fm/api
5. N/A. (2022, December 29). *Billboard Top 100 Songs of every year*. Billboard Top 100. Retrieved April 18, 2023, from http://billboardtop100of.com/
6. Spotify. (n.d.). *Web api*. Web API | Spotify for Developers. Retrieved April 18, 2023, from https://developer.spotify.com/documentation/web-api

Figures

Top 10 Artist with the most appearances

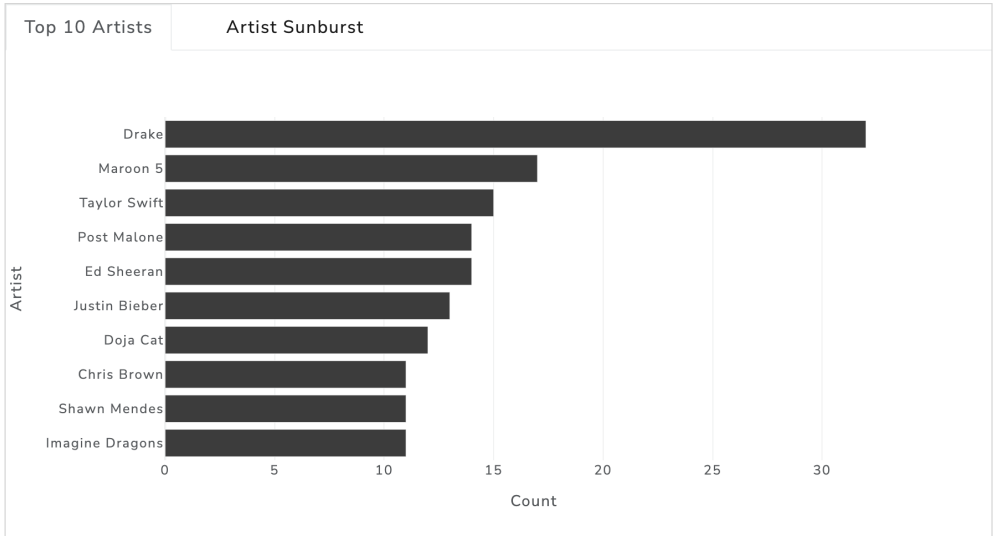

Figure 1 - Vertical bar plot of top 10 Artists with the most appearances within the selected year range. From 2012 to 2022, Drake is tremendously popular with more than 30 times appearance and Maroon 5, Taylor Swift, Post Malone, and others follow him behind around 10-15 counts.

Top 10 Artists with most songs by genre

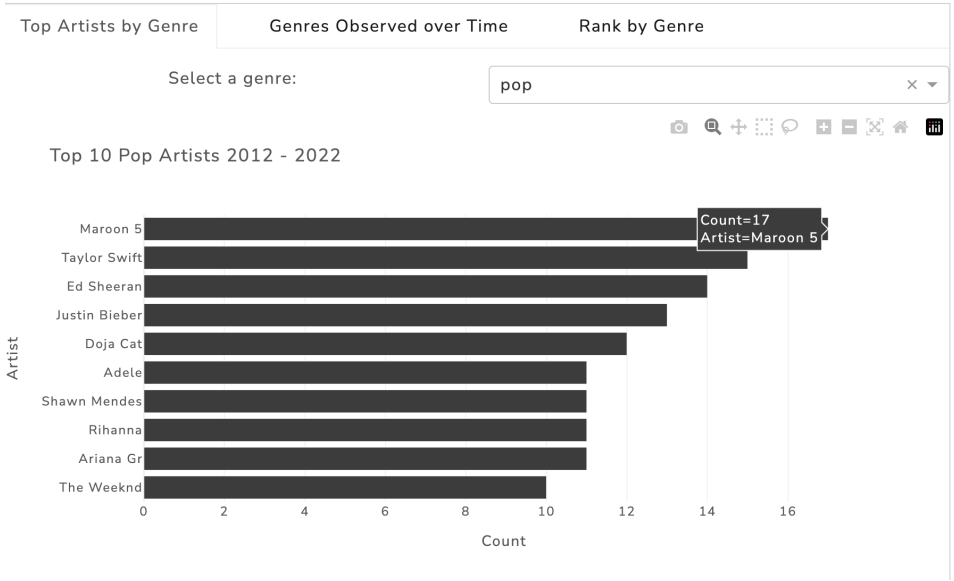

Figure 2 - Vertical bar plot of top 10 artists with the most songs by genre between 2012 to 2022. Genre can be changed in the drop-down menu on the top right corner of the visualization. Maroon 5 has been the most popular artist for pop music between 2012 and 2022 with 17 pop songs, with Taylor Swift being a close runner up with 15 songs in the Top 100 charts.

Sankey from artist to year to danceability

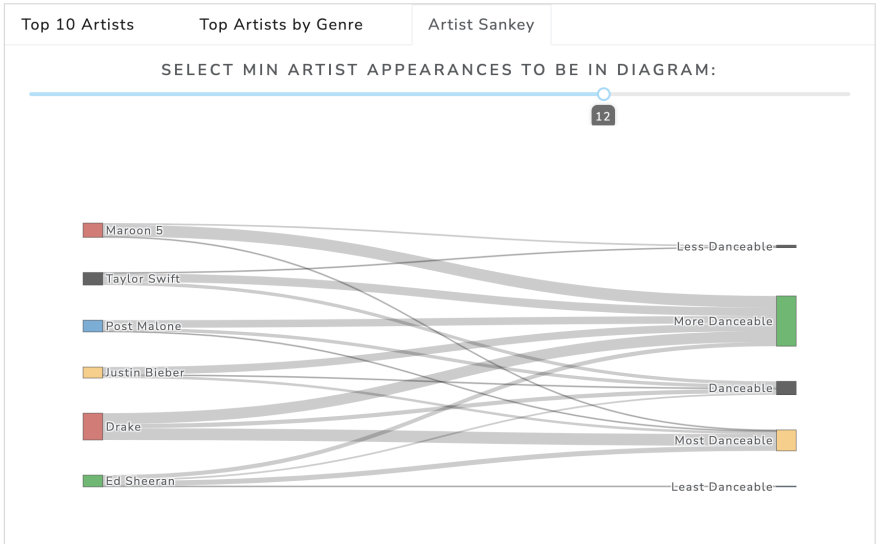

Figure 3 - Sankey diagram that visualizes the relationships between artists, year, and danceability between 2012 to 2022. Drake has the most diverse portfolio out of all artists, but Maroon 5 actually had the the the highest amount of "more danceable" songs with 14 compared to Drake's 13. The cutoff for quantity is static at 12 songs. 2019 had the highest number of danceable songs (avoiding less and least danceable song types).

Sentiment Over Time

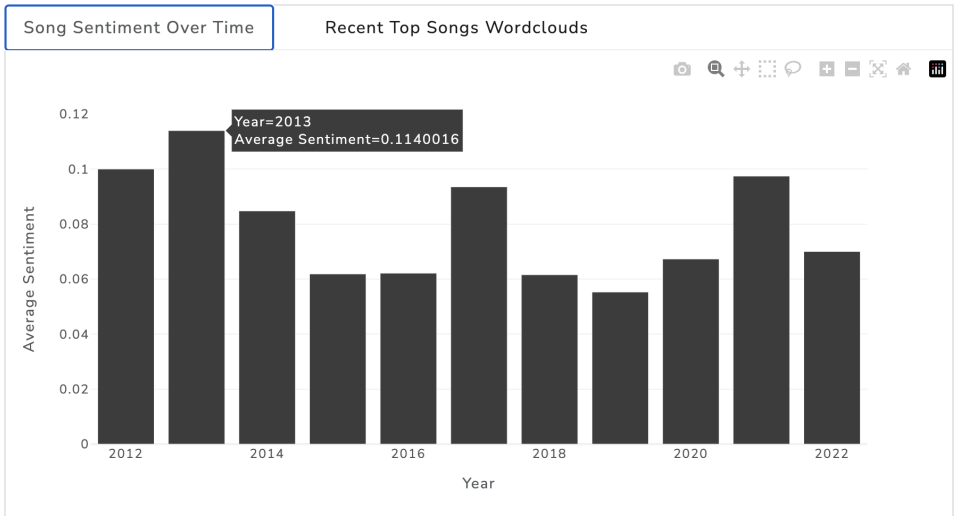

Figure 4 - Bar plot of average sentiment score over time from 2012 to 2022. It allows the user to see the general trend of sentiment positivity/negativity each year within the year range. From 2012 to 2022, the positive songs generally ranked in Billboard Top100 songs.

Sentiment Color Palette

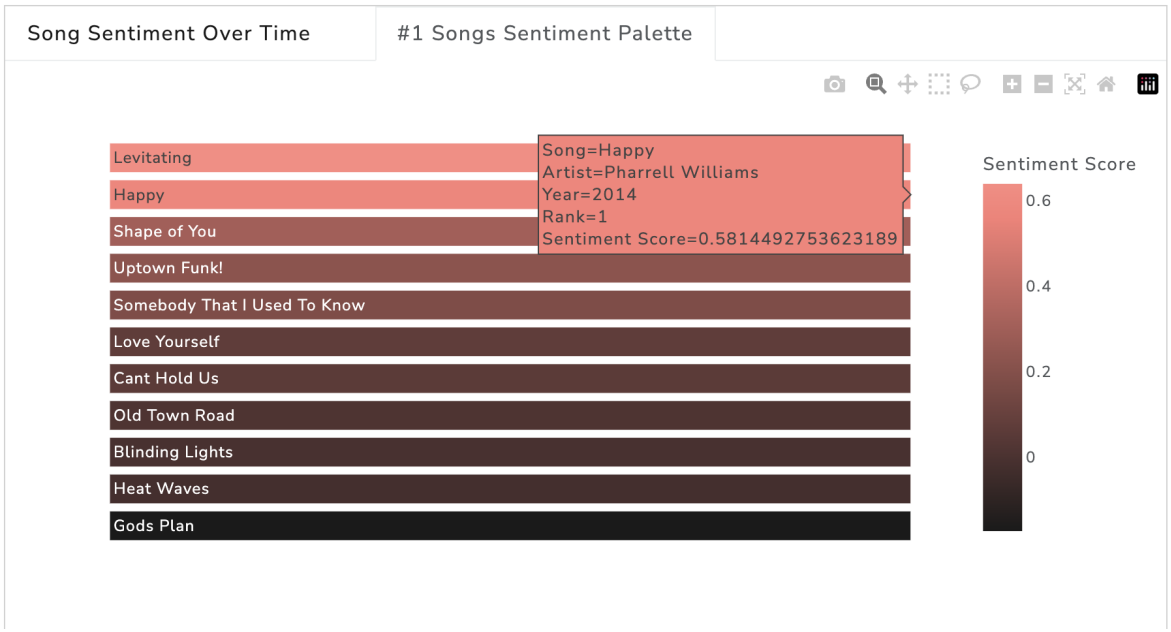

Figure 5- A color palette visualization that sorts the number one songs between 2012 to 2022 by their sentiment scores to see which songs are the most positive. The darker colored songs are affiliated with more negative lyric sentiments while the brighter songs are relatively more positive.

Sankey from genre to year to artist

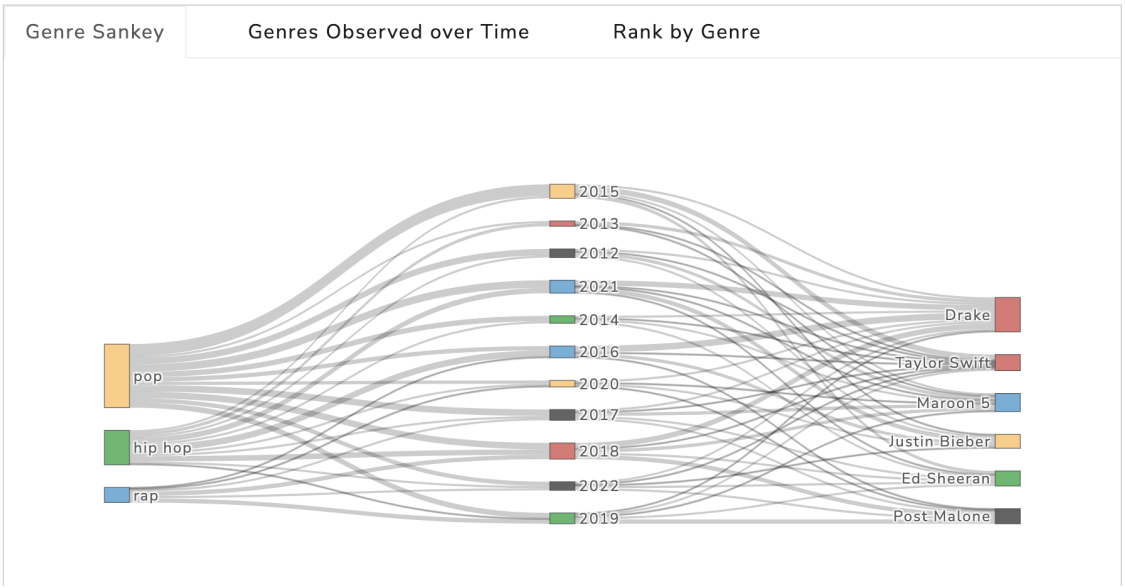

Figure 6 - Sankey diagrame that compares the most popular genres between 2012 to 2022, the quantity of songs in the year, and the artist affiliated with the song. The cutoff for quantity is static at 12 songs. Pop is the most common across all years, followed by hip hop and rap. For artists, Drake had the most diverse portfolio among all artists between 2012 and 2022.

Genre Over Time

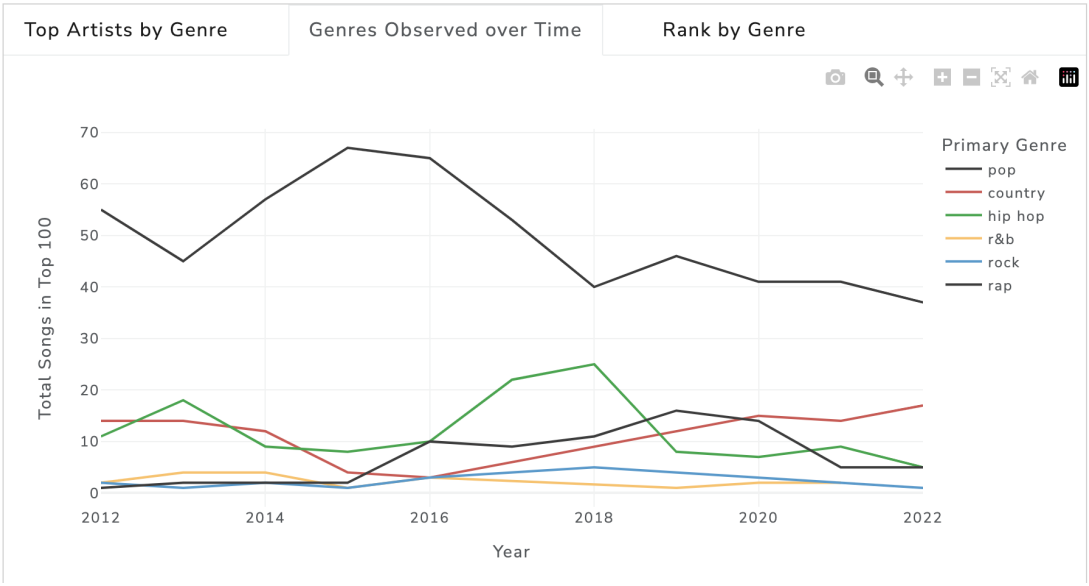

Figure 7 - Genres observed over time from 2012 to 2022. Pop has decreased since 2015, while country music has risen steadily since 2016. Hip hop and rap have decreased since 2018-2019. Rock and R&B have consistently been the least popular genres since 2012. Unknown genres were deselected for clean viewing.

Rank by Genre

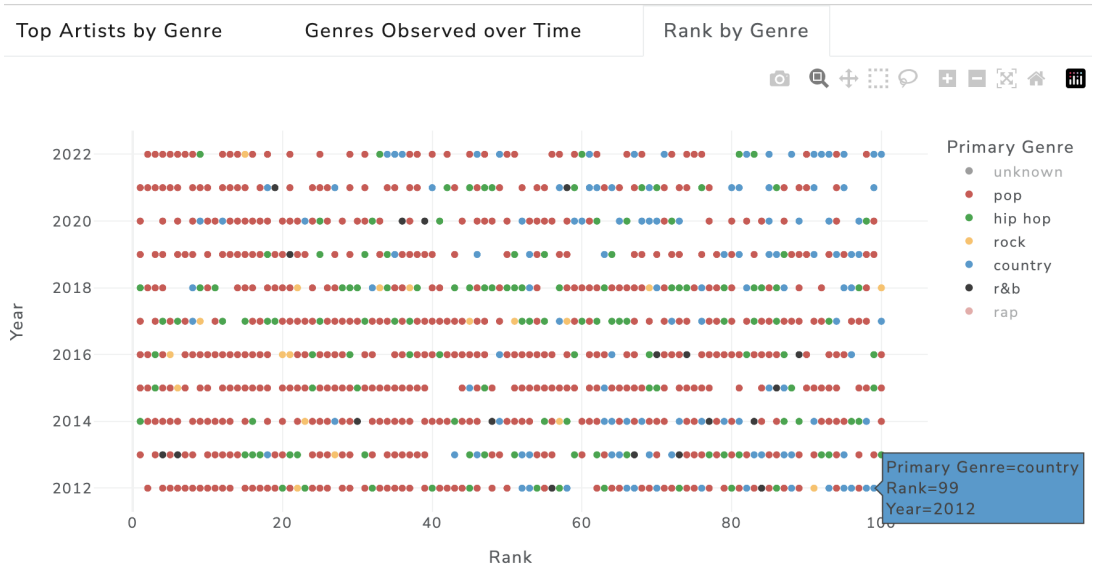

Figure 8 - Scatterplot depicting the distribution of genre over time, with the first rank being on the left, and descending in ranking as the x axis increases. Unknown genres and rap have been removed to accentuate the location of pop songs in rankings. This figure enables users to see the general popularity of genres within the top 100 songs for each year, and where they are

located in terms of ranking. For years 2019 to 2022, pop songs have dominated the majority of the top 20 songs, and have remained prevalent from 2012 to 2016.

Recommendation system

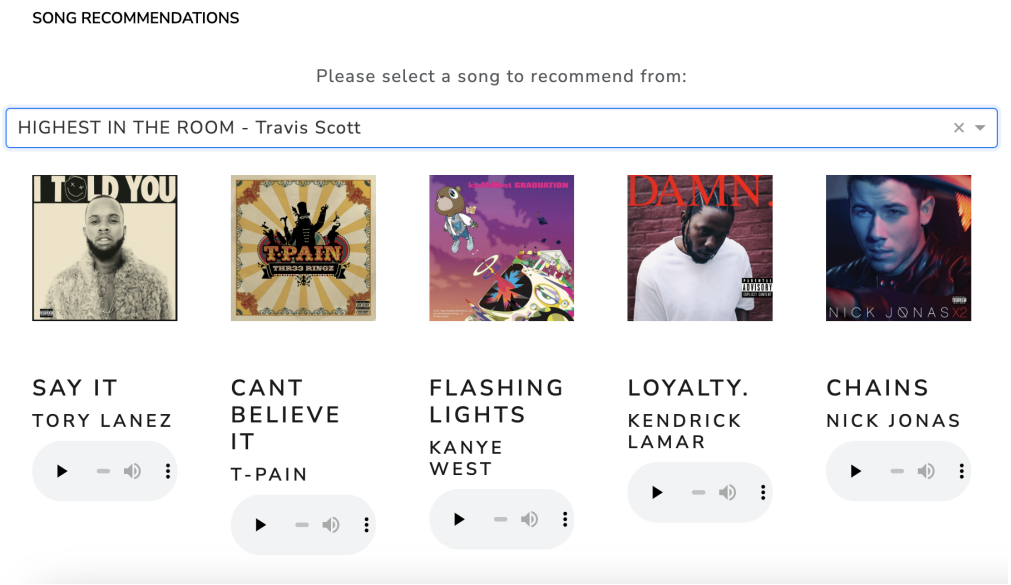

Figure 9 - Song recommendation system based off of all songs from 2000 to 2022. The recommendation system implements the K Nearest Neighbor algorithm to choose songs that closely fit the selected song's characteristics, such as valence, danceability, popularity, etc. Travis Scott's "Highest in the Room" is a rap song, and the recommendation system proposes 4 rap/hip hop songs, with the exception of Nick Jonas.