Assessment 2: Data Visualization.

Most Stream Spotify 2023.

03/11/2023.

Student: Huynh Quang Minh NGUYEN

Student ID: 46863885

Lecturer: Geng Sun

Class: Friday 5 pm - 8 pm

Unit: Data and Visualization for Business (BUSA8090)

1. Describe the four data analytics process for data governance. List the features of R/tableau that can help you to provide visualisation according to data governance principles.

To make data stewardship and governance simpler, we will divide the DGI Data Governance Framework (2023) into four phases: "Why - What & How - Who - When" and emphasise the crucial business activities.

- Discovery process: when implementing the DGI data governance paradigm in a value-driven data governance programme, the first question to ask should be "WHY". This approach uses a mission and vision statement that outlines the advantages of data governance for a business. This should, at the very least, have the backing of senior management and be linked to the strategic objectives of the organisation.
- Define Process will be used to illustrate the "What & How" problem. As opposed to

  + What: The program's short- and long-term goals, success standards, and assessment methods. This often necessitates addressing the main issues that surface in various business units. This has to be followed by the finance department and any other pertinent line managers.

  + How: Definitions of data, data standards, data policies, and the process of converting data rules into business rules (preferably in the form of a business lexicon). We will now talk about the data assets that identify the important business entities needed to meet the goals of the organisation.

- Apply process “WHO” starts with the support of the cross-functional data governance activities and procedures, establishing by the Governance Office or Team. It collects metrics and success measures and reports them to data stakeholders in addition to providing constant communication, information access, record-keeping, training, and support to stakeholders. Data stewards will now be essential to maintaining data rules and fixing many issues before they become major roadblocks.

- The last process will be Measure and Monitor processes “WHEN” in which a set of standardised, documented, and repeatable procedures must be put into place together with the necessary number of auxiliary technologies. In the end, how effectively processes are connected will determine how successful data governance framework is and how far we can take data governance maturity.

b, The feature of R that help to provide visualization according to data governance principles

A collection of rules called data governance principles aids businesses in efficiently managing their data. Initiatives related to data governance are framed by the guiding principles (Data Governance Institute, 2023). Among these are data standardisation, auditing, change management, accountability, transparency, and data integrity preservation.

Numerous characteristics of R help to reinforce the concepts of data governance. The capacity to generate data dictionaries is one of the most crucial characteristics. A dataset's format, content, and organisation are all described in a data dictionary. It offers a thorough summary of the data along with variable names, descriptions, data types, and other pertinent information. This might help guarantee the accuracy, consistency, and timeliness of the data.

Numerous built-in tools for data visualisation are available in R. “Ggplot2” is one of the most widely used packages. We may build a variety of visualisations using this strong and adaptable tool, such as scatter plots, line charts, bar charts, and more. The foundation of the system, The Grammar of Graphics, offers a strong graphics paradigm that facilitates the creation of intricate, multi-layered images. “Pacman” is an additional tool that might be helpful for data visualisation. It is a package management tool that makes loading and installing R packages easier. When using a range of technologies for data analysis and visualisation, this might be helpful.

Version control is another R method that might be useful for data governance. We can monitor the changes made to data over time using version control. This might guarantee that the data is auditable and that updates are done in a transparent, controlled way.

Lastly, R offers a plethora of tools for data cleansing and verification. We can be sure that data is precise, comprehensive, and consistent when we use these solutions. They could also assist us in finding mistakes and discrepancies in ourdata.

2. Ethics is an important aspect of Data visualisation, discuss and apply any three ethics values that I have maintained.

Ethical dilemmas are common in the data world because data collection, analysis, transmission, and usage can have a substantial impact on people's and organisations' ability to live well (Vallor, S., 2018). The following three moral guidelines should be kept in mind while creating data visualisations:

1. Privacy: one of the most crucial ethical considerations while making data visualisations is privacy. Respect the privacy of everyone whose information has been acquired, and never release personal information without consent. It's likely that most of us are ignorant of the degree to which ordinary data practises may expose our personal information given the amount of personal data that individuals are already creating (Vallor, S., 2018). Unauthorised access and information reuse have the potential to harm people, make it harder for them to maintain a condition of restricted access to their personal data, and eventually harm an organization's reputation in its interactions with shareholders, customers, and government agencies, claim (Culnan, M.J. and Williams, C.C., 2009). Mandinach, Parton, Gummer, and Anderson (2015) contend that it is critical to comprehend how to use data responsibly while maintaining secrecy and privacy.
2. Transparency: this is particularly crucial when making data visualisations to ensure complete openness and respect for the data sources. Data practises may change social openness in both directions, however for now, reducing transparency is riskier for the two reasons mentioned below (Vallor, S., 2018). The quantity and complexity of data that is now accessible is the first issue that poses a concern. The second problem is that trade secrets and proprietary technology are often used to hide data practises (Vallor, S., 2018). Therefore, we may express our worries about big data considering specific moral norms more effectively by using practical ethical theorie (M.J. Quinn, 2018).

3. Honesty is a crucial component of data ethics. Experts in ethical data never intentionally change, distort, or obstruct comprehension; instead, they always provide information in the most objective and accurate way possible. Even in situations when it appears improbable that they would be caught, honest individuals never steal, cheat, or lie. Participants in the collection, maintenance, use, and reporting of education data include volunteers, vendors, employees, and appointees. They all really need to demonstrate professionalism, ethics, and honesty on a regular basis. These qualities serve as the cornerstone of moral conduct (NCES, 2010).

3. Discuss Task A visualisation and provide sufficient justification on the prediction.

Visualization 1: Music analysis in Audio Features

Question: What is the pattern in audio features to help better understand trends and preferences in popular songs.

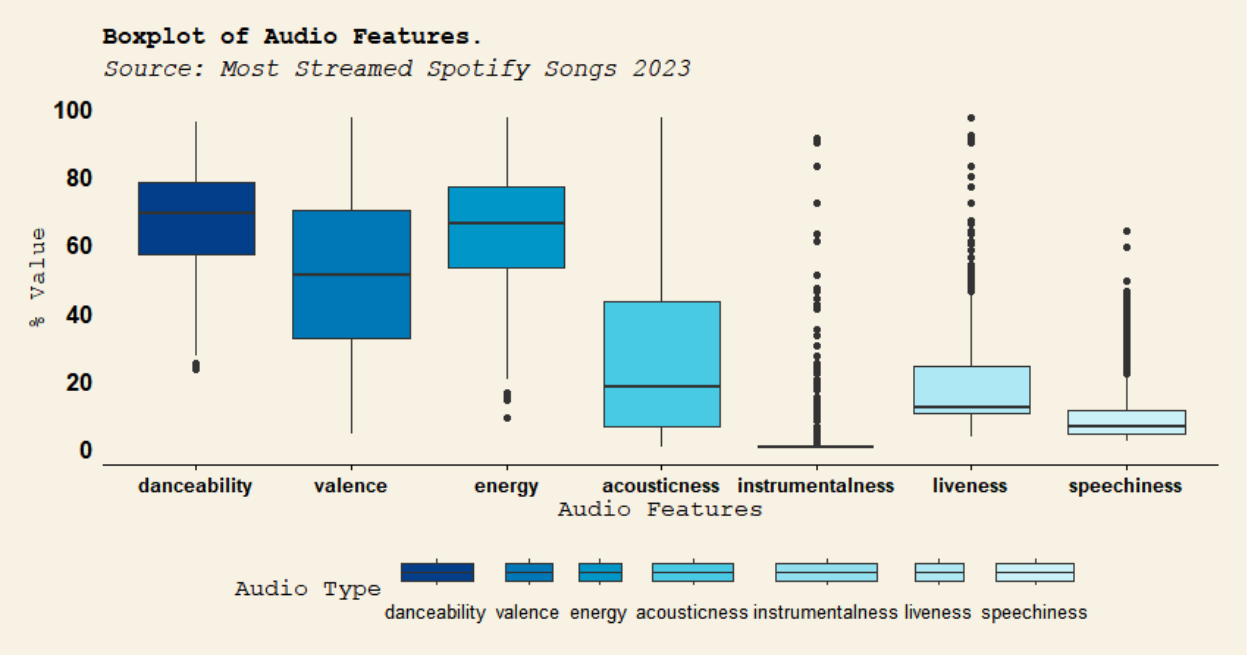

Figure 1: Audio Features

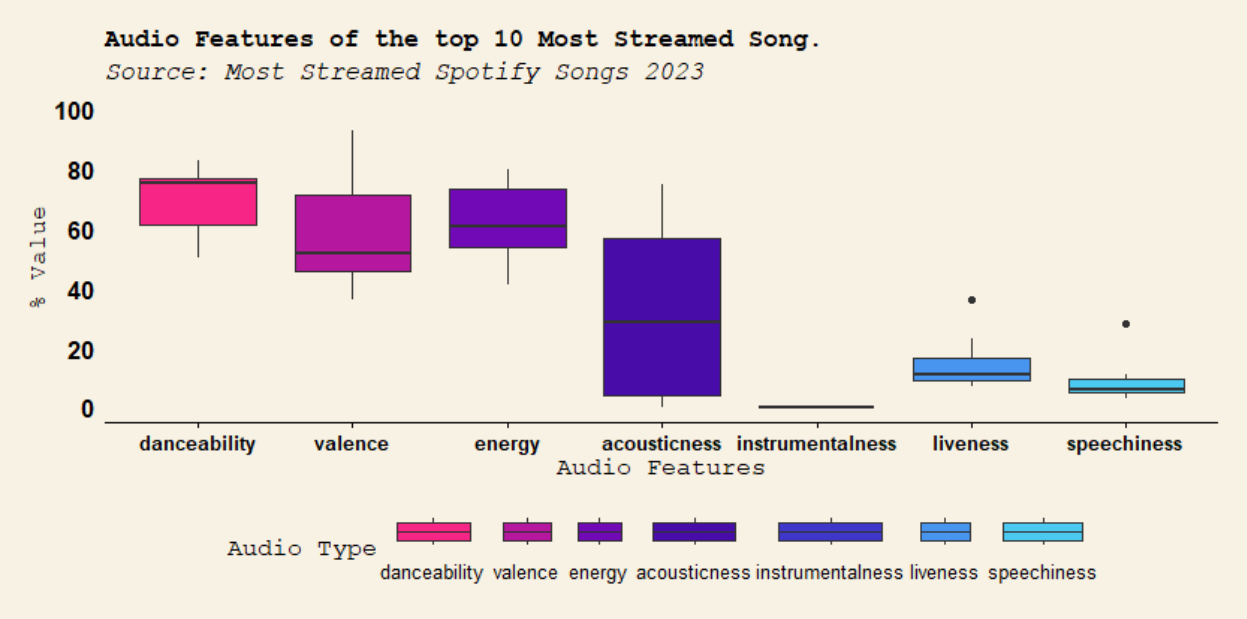

Figure 2: Audio Features of the top 10 Most Streamed Song

Interpret the result

Based on the box-plot analysis, there appear to have useful insights through different audio features and the mean values of the audio feature across songs in comparison with the top 10 most streamed songs is similar to each other.

- Danceability: Since most songs have high danceability scores with the average danceability score is around 67 while the top 10 song is 69.7 (on a scale of 0 to 100), it suggests that listeners prefer songs they can dance to.
- Valence: The wide range of valence scores with the average valence (a measure of musical positiveness) is around 51 vs 59.2, indicates that listeners have diverse tastes when it comes to the mood of the music.
- Energy: As most songs have high energy with the average energy score is around 64, it might be a good idea to focus on promoting high-energy music.
- Acousticness and Instrumentalness: The low scores in these areas with the mean of 27 suggest that listeners prefer songs with vocals and less acoustic sound.
- Liveness: The low liveness scores with 18% vs 14.7% on average suggest that listeners prefer studio recordings over live recordings.

- Speechiness: The low speechiness scores around 10% on average indicate that listeners prefer music over spoken words. However, the presence of high speechiness songs suggests there might be a market for podcasts or audiobooks on the platform.

Recommendations for Sony Music's CMO

1. Focus on Danceable and Energetic Music: Since danceability and energy are common features of popular songs, it would be beneficial to focus on promoting songs and artists that produce danceable and energetic music. Therefore, promoting songs with high danceability could potentially increase streams. It might be beneficial to segment the audience based on their preference for positive or negative songs and target them with personalized recommendations.
2. Promote Vocal Tracks: As instrumental tracks are less popular, the focus should be on promoting tracks with vocals. However, the outliers indicate there is a niche audience for acoustic and instrumental music. Identifying and targeting this audience could be a potential growth area.
3. Studio Recordings Over Live Performances: Since songs with low liveness scores are more popular, it would be beneficial to invest more in studio recordings as opposed to live performances. However, offering exclusive live recordings could be a way to differentiate and attract a niche audience.
4. Mood Doesn't Matter as Much: Since the valence scores vary widely among popular songs, it suggests that the mood of the song doesn't significantly impact its popularity. Therefore, the company can diversify its portfolio by promoting songs of various moods. It might be beneficial to segment the audience based on their preference for positive or negative songs and target them with personalized recommendations.

Visualization 2: Music analysis in BPM

Question: What is the BPM that customer interested most in. Therefore, help the CMO to listen to the trends of the market and directing the market research efforts of the company.

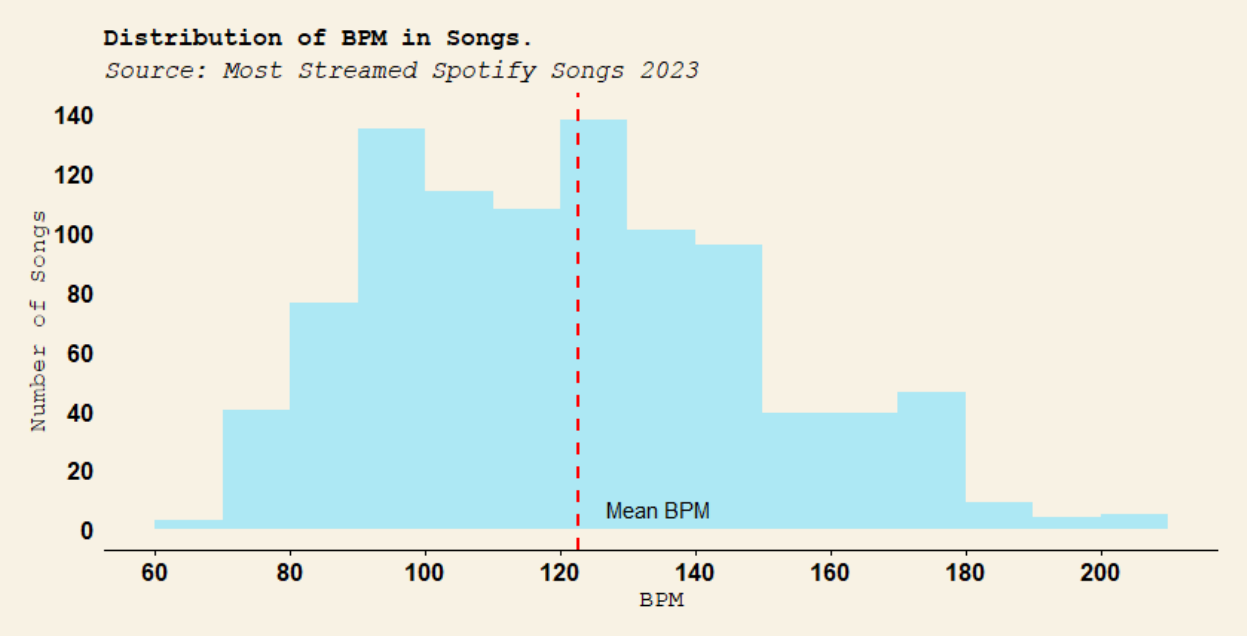

Figure 3: Distribution of BPM in Songs

Interpret the result:

Analyzing the histogram of BPM (Beats Per Minute), we can derive several insights:

- Distribution of BPM: The histogram provides a visual representation of the distribution of BPM in the songs. The peak of the histogram indicates the most common BPM range for songs, which is around 120-140. The distribution is likely skewed to the left, it indicates that there are more songs with lower BPM.
- Average BPM: The red dashed line represents the average BPM of the songs. This can be used as a benchmark to understand what BPM is considered 'typical' for a song.

Recommendations for Sony Music's CMO

1. Target the Popular BPM Range: A certain BPM range is found to be around 90 – 130 with more popular among listeners, the CMO could focus on promoting songs within this BPM range to attract more listeners.
2. Explore Different BPM Ranges: As the distribution of BPM is likely skewed to the left, it suggests that there's a lack of songs that have high BPM. The CMO could consider exploring these BPM ranges to provide more variety to listeners and potentially attract a new audience.

Visualization 3: Cross-platform analysis

Question: investigate how songs perform across different streaming services. Therefore, can help Sony Music's CMO to define marketing strategies to support the company's overall strategies and objectives.

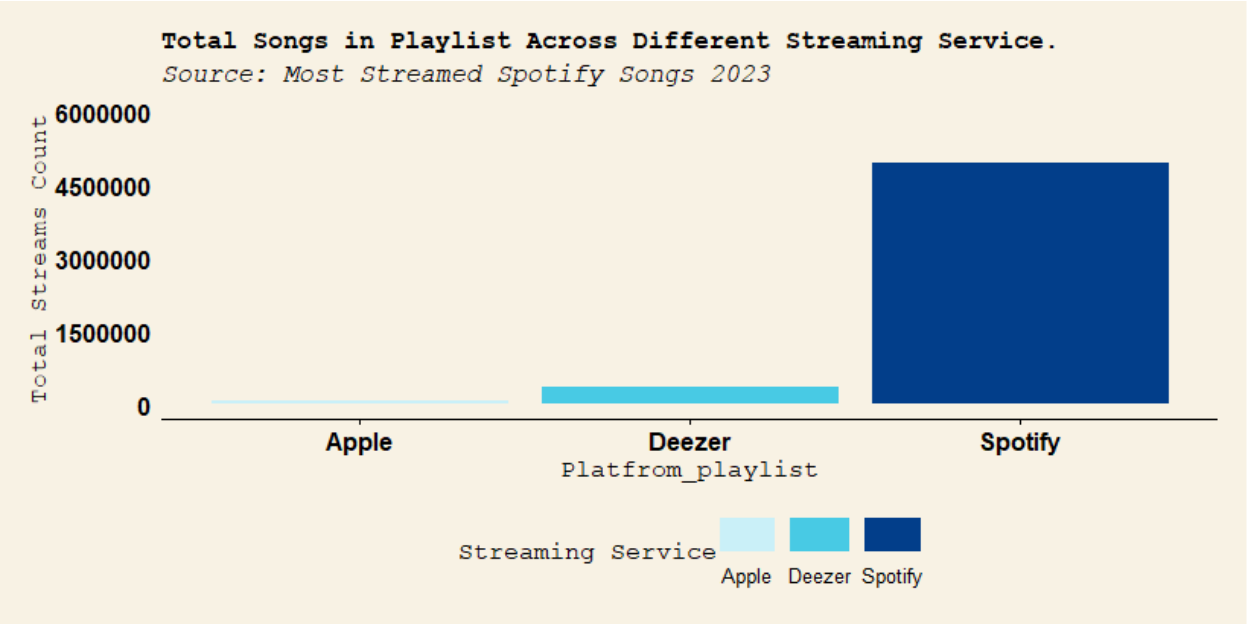

Figure 4: Total Songs in Playlist across different Streaming services.

Interpret the result:

- The bar chart shows the total count of songs in playlists across three different streaming services: Spotify, Apple, and Deezer. From the chart, we can observe that Spotify has the highest count of songs in playlists, followed by Deezer and then Apple. This could suggest that Spotify's playlists are more diverse or that Spotify users are more active in creating and sharing playlists.

Having known Spotify and Deezer playlists has the highest playlist, we could further investigate the relationship between the number of songs in playlists across these 2 platforms to know whether they have positive or negative relationship. This could provide additional insights into user behavior and preferences on these platforms.

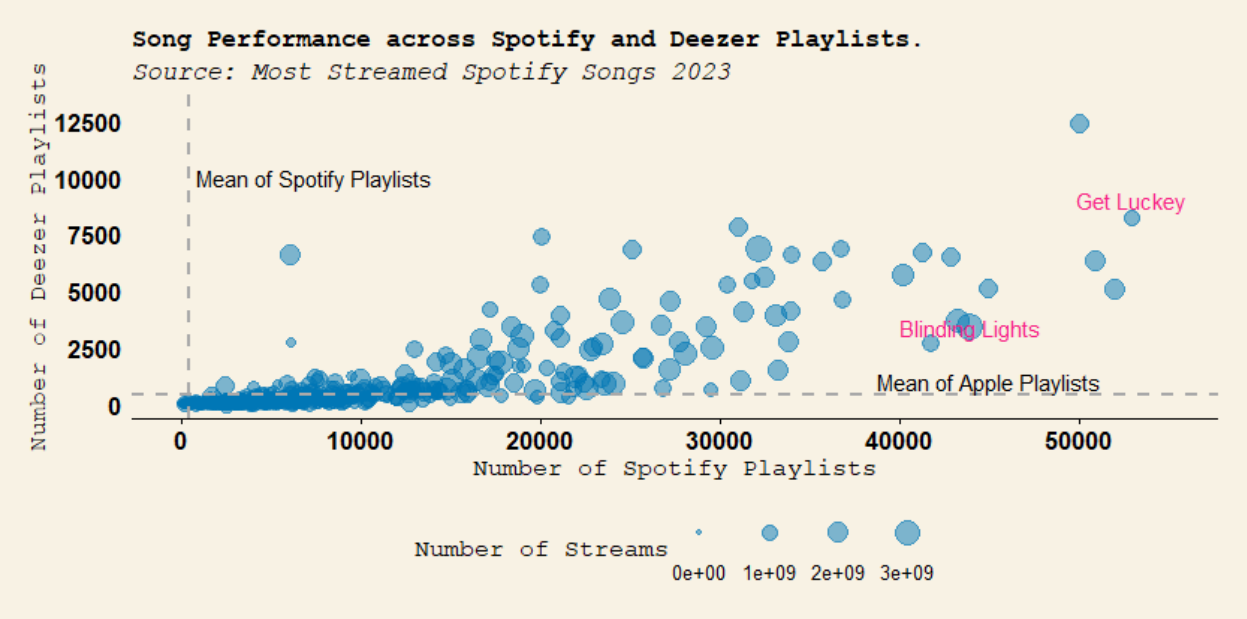

Figure 5: Song Performance across Spotify and Deezer Playlists.

Interpret the result:

- The scatterplot shows the relationship between the number of Spotify playlists and Deezer playlists a song is included in. From this graph, we can clearly observe that: there appears to have a positive correlation between the number of Spotify playlists and Deezer playlists a song is included in, it suggests that songs popular on one platform are likely to be popular on the other as well.

Recommendations for Sony Music's CMO

1. Focus on Popular Songs: Most of the points in the scatter plot are concentrated towards the left side of the plot and fewer points are spread out towards the right, this could suggest a right-skewed distribution for both Spotify and Deezer playlists. This would mean that most songs are present in a smaller number of playlists, but there are a few songs that are present in many playlists. Therefore, one of the most effective strategies is to focus promotional efforts on a small number of popular songs could be an effective strategy.

2. Playlist Inclusion: Work on strategies to get songs included in more playlists on both platforms. Being in a playlist increases a song's exposure and can lead to more streams.
3. Partnerships and Collaborations: Spotify will be the best choice due to its larger playlist count when the company is considering partnerships or collaborations with streaming services. Also, if a song is popular on one platform, consider promoting it on the other platform as well to leverage its popularity as the scatterplot shows the positive correlation.
4. User Engagement: The company could investigate why Spotify has a higher song count in playlists. Understanding the factors that contribute to this could help the company increase user engagement on other platforms.

Visualization 4: Artist Impact

Question: How artist involvement and attributes relate to a song's success. Help the CMO to liaise with other departments to guide a unified approach to customer service, distribution etc. that meets market demands, collaborates with cross-functional teams to create and implement integrated marketing campaigns for artists.

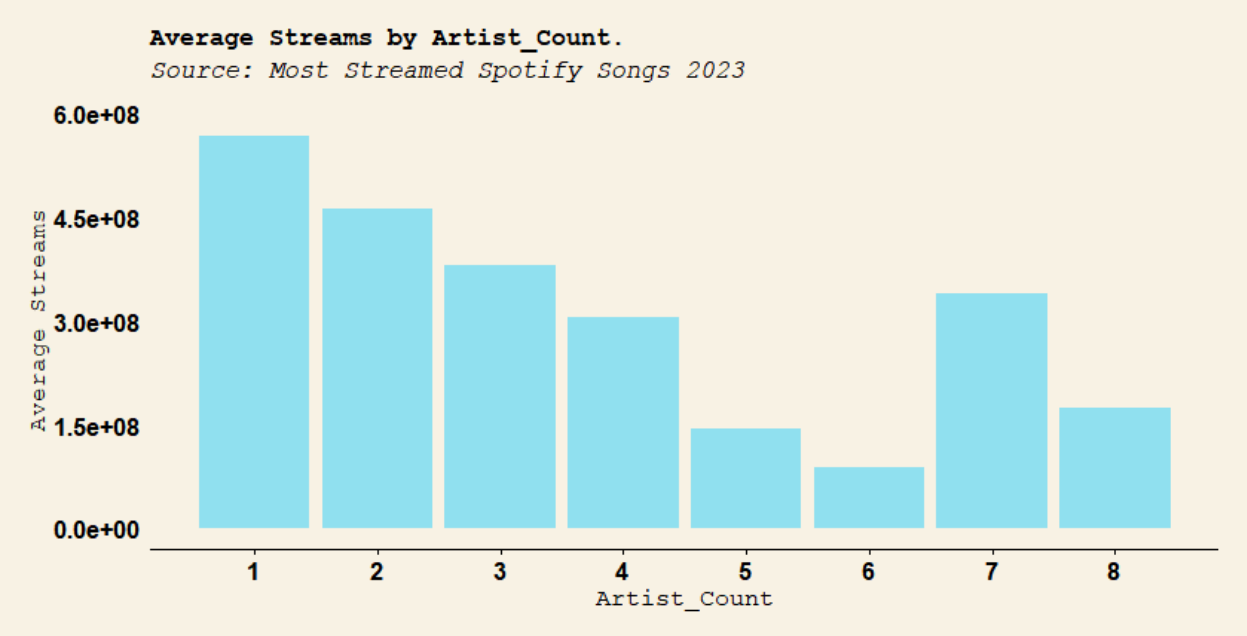

Figure 6: Average Streams by Artist Count.

A combined histogram and line chart showing the artist count with their total streams and Spotify playlist count. The histogram represents the total streams, and the line chart represents the Spotify playlist count. This chart can help analyze how the number of artists involved in a track relates to its success in terms of streams.

Interpret the result

- Tracks with a higher artist count tend to have more streams and are included in more Spotify playlists. This could be due to the combined fan base of multiple artists leading to more exposure and streams.
- There is a peak in the total streams and Spotify playlist count at a certain artist count. This suggests that there is an optimal number of artists to involve in a track to maximize its success.

Recommendations for Sony Music's CMO

1. Collaborations are Key and become a new trend. Therefore, Sony Music should encourage collaborations between artists. The combined fan base can lead to more streams and playlist inclusions.
2. Optimal Artist Count: The optimal number of artists to be involved in a track is supposed to be from 1 – 3 artists involve in 1 song. Too many artists might dilute the impact of individual artists, while too few might not provide enough exposure.

Visualization 5: Artist Popularity

Question: How can an artist increase their total streams? Does being included in more Spotify playlists necessarily lead to more streams?

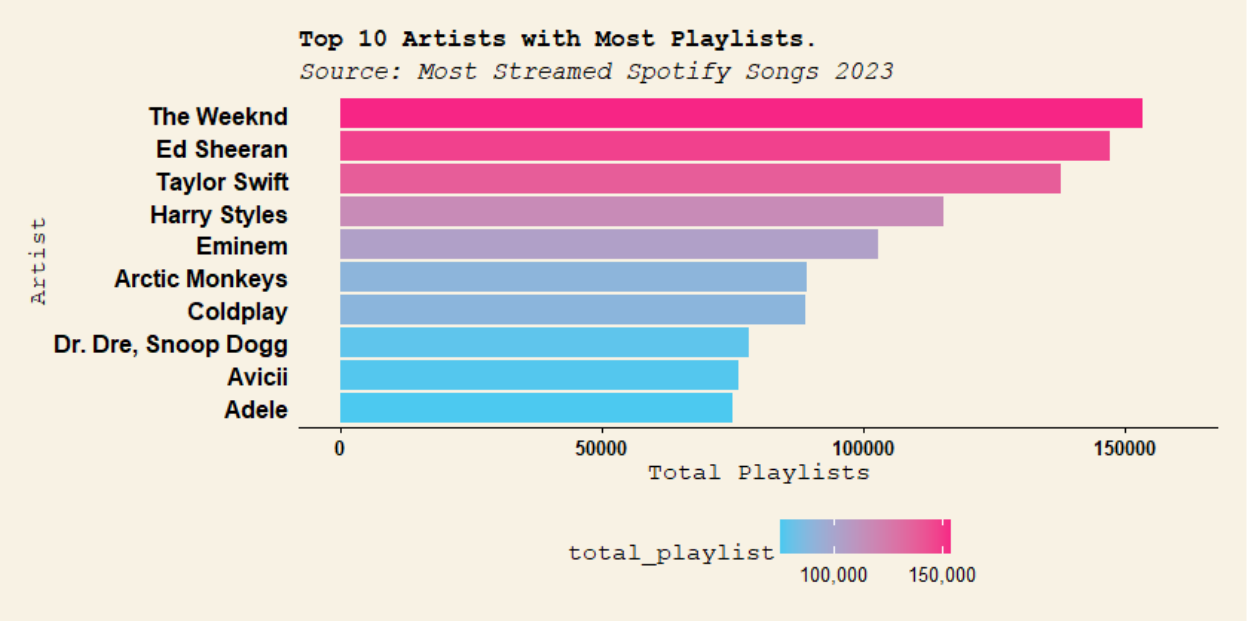

Figure 7: Top 10 Artists with Most Playlists.

The chart displays the top 10 artists who are featured in the most playlists. This chart is useful for understanding which artists' music is most included in playlists as it could indicate a high level of popularity or broad appeal, as playlists often cater to a wide range of musical tastes.

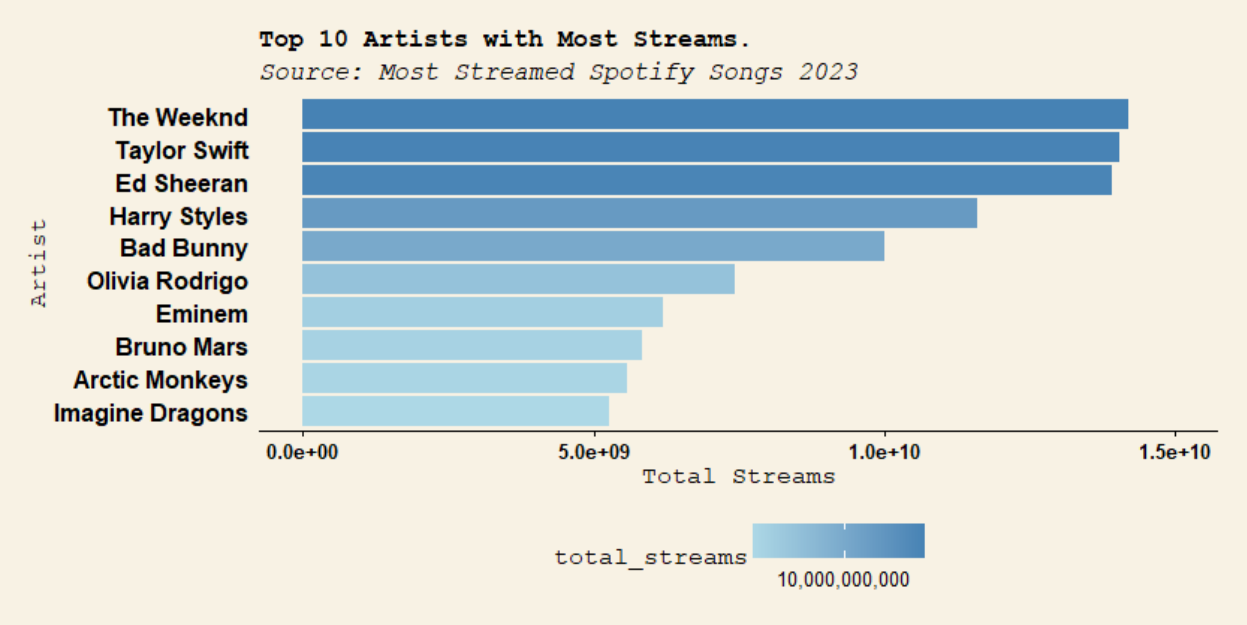

Figure 8: Top 10 Artist with Most Streams.

The chart displays the top 10 artists with the most streams. This chart is useful for understanding which artists are the most popular or have the most listened to music, based on the number of streams.

Having known the top 10 artists with their most streams and playlist. We then conducted the scatterplot to show the relationship between total playlists and total streams.

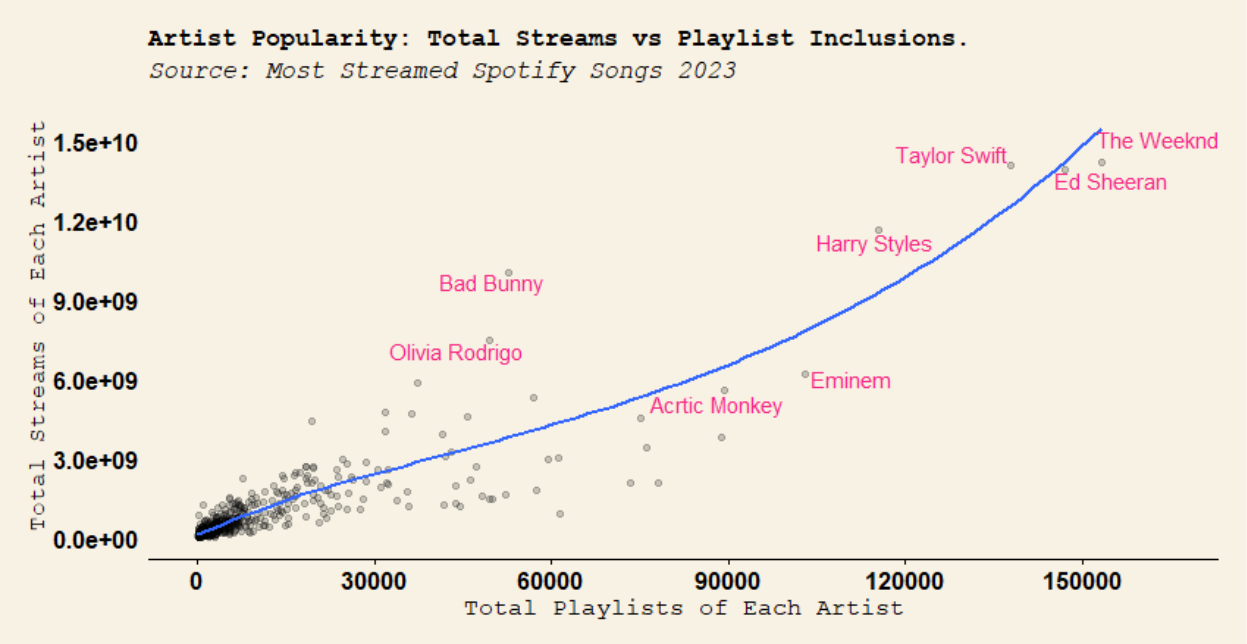

Figure 9: Artist Popularity with their Total Streams and Playlist Inclusion

Interpret the chart:

- Positive Correlation: There appears to be a positive correlation between the number of playlists a song is and the number of streams the song has. This suggests that artists with a higher number of songs in Spotify playlists tend to have more streams. This could be because being featured in a playlist increases the visibility and accessibility of an artist's songs, leading to more streams.
- There are some artists with fewer playlist inclusions like Bad Bunny, Olivia Rodrigo but a high number of streams. This could be due to factors like the artist's existing popularity, the quality of their songs, or successful marketing strategies outside of Spotify.
- Density of Points: The density of points can give us an idea about the commonality of songs in terms of their presence in playlists and their streams. There are more points in a

0 – 10000 Playlist Area of the plot, it means that there are more songs with those characteristics.

- Trend Line: The trend line (or regression line) represents the average relationship between “playlist” and 'streams'. It shows that on average, as the number of playlists a song is in increases, so does the number of streams.

Recommendations for Sony Music’s CMO

1. Promotion of Songs: Promote songs to the audience and encourage them to add songs to their personal playlists. This could increase the visibility and streams of songs.
2. Look into the artists who have high streams despite fewer playlist inclusions. Analyze their marketing strategies and see if there are any successful tactics that can be adopted.
3. Identifying Popular Artists: These artists clearly resonate with listeners, as their music is being included in many playlists. Understanding why these artists are popular could help in developing strategies for other artists.
4. Partnership Opportunities: These artists could be good candidates for partnerships or collaborations, as their music is widely listened to.
5. Marketing Strategies: The popularity of these artists on playlists could be leveraged in marketing campaigns. For example, a campaign could highlight the fact that an artist is featured in a large number of playlists.

Reference

1. Culnan, M.J. and Williams, C.C., 2009. How ethics can enhance organizational privacy: lessons from the choicepoint and TJX data breaches. MIS quarterly, pp.673-687.
2. DGI Data Governance Framework (2023) The Data Governance Institute. Available at: https://datagovernance.com/the-dgi-data-governance-framework/.
3. E.B. Mandinach, B.M. Parton, E.S. Gummer, R. Anderson, Ethical and appropriate data use requires data literacy, Phi Delta Kappan 96 (5) (2015) 25e28.
4. Goals and principles for Data Governance (2023) The Data Governance Institute. Available at: https://datagovernance.com/goals-and-principles-for-data-governance/ (Accessed: 03 November 2023).
5. M.J. Quinn, Ethics for the Information Age, Pearson, Hoboken, New Jersey, 2017.
6. The Forum Guide to Data Ethics (2010) National Center for Education Statistics (NCES) Home Page, a part of the U.S. Department of Education. Available at: https://nces.ed.gov/pubs2010/dataethics/honesty.asp.
7. Vallor, S., 2018. An introduction to data ethics. Course module.) Santa Clara, CA: Markkula Center for Applied Ethics.