Part A - Practical Work

1. Introduction

The confluence of data analytics and the music industry in the digital age has revealed deep insights into customer behavior and preferences, revolutionizing the creation, distribution, and consumption of material. An unparalleled amount of data has been made available by the rise of streaming services, providing detailed insights about listener engagement across multiple platforms. Record labels, musicians, and marketing experts can all benefit from the strategic insights this report provides by using R, a statistical computing and graphics language, to analyze and interpret this data.

We have created a number of visualizations that decipher intricate patterns in streaming data using R's powerful packages, like ggplot2 and dplyr. To illustrate the various ratios of energy, valence, acousticness, and other musical elements that appeal to listeners, the distribution of song qualities for the top 50 songs by streaming is displayed. Large datasets are simplified into easily understood formats by these visual representations, which also make it easier to compare investigations across many metrics and temporal dimensions. This allows for the identification of both permanent patterns in music popularity and transient fads.

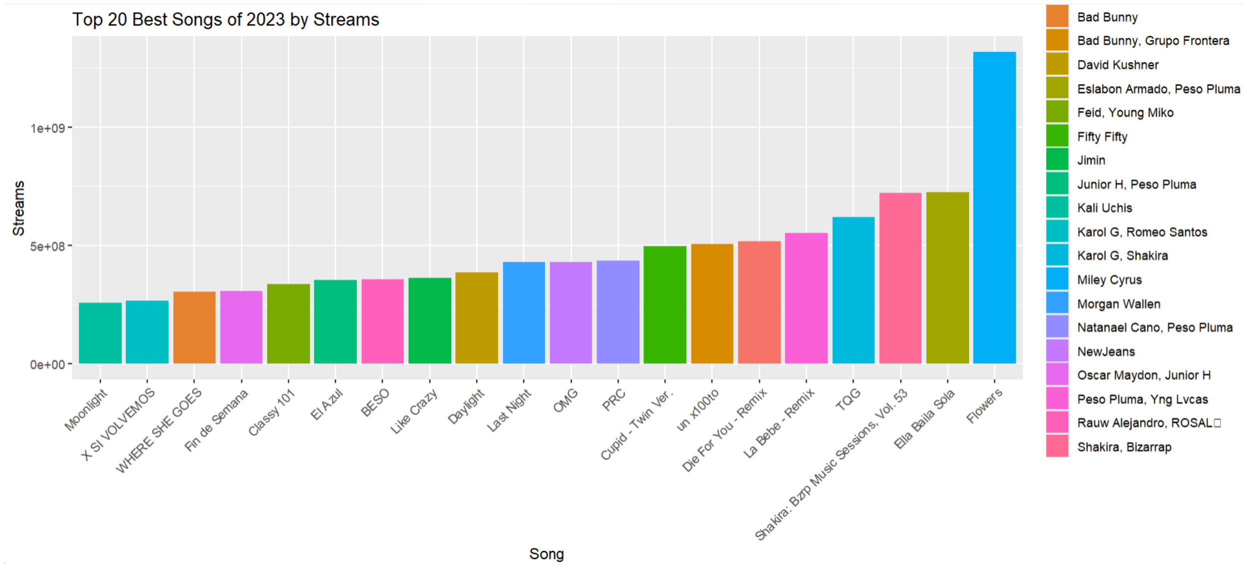
To highlight changes in musical tastes and inclinations, the analytical journey starts with a graphic representation of the top songs from both the current year and the history of music. The effects of artist collaborations on streaming are examined in further analyses, which are succeeded by a cluster analysis that organizes songs logically according to their auditory characteristics. Through the analysis of new trends and listener preferences, this scientific approach not only determines the qualities shared by the most streamed songs, but it also makes predictions about future hits.

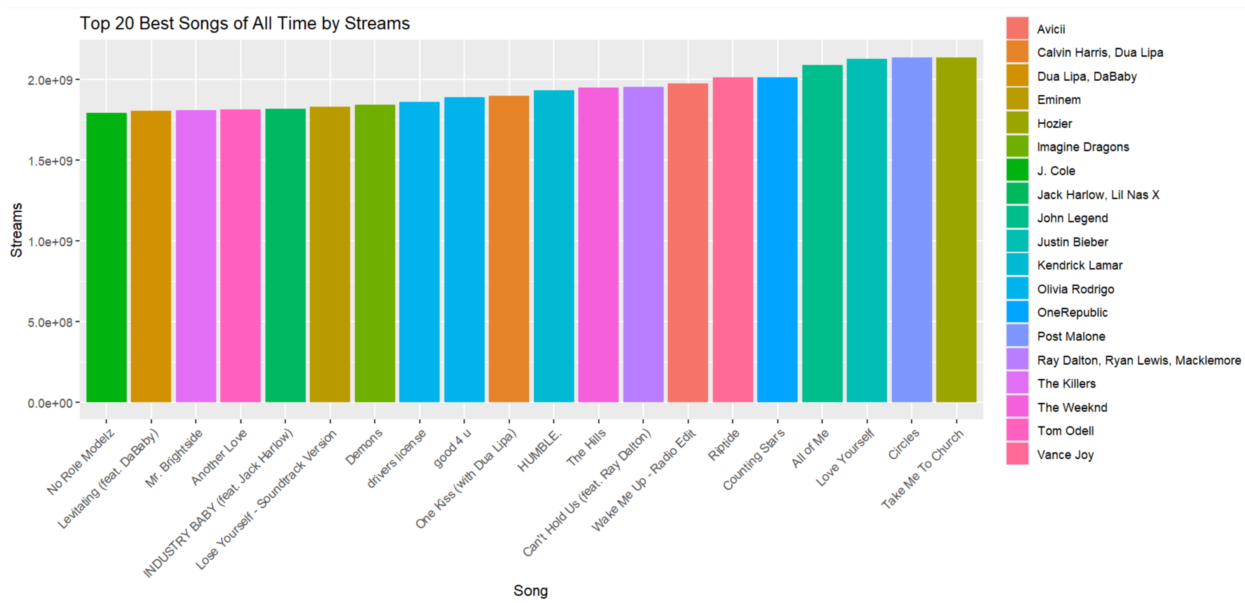
By recommending potential future routes for the sector, predictive analytics contributes even more to our understanding. For example, the consistency of several sound features in the most streamed songs points to a formula that future releases may find success with fans. Additionally, strategic insights into optimizing cross-platform performance, boosting promotional efforts, and expanding audience reach can be gained by analyzing correlations between various streaming platforms and KPIs.

The ultimate goal of this research is to provide professionals in the music industry with useful information gleaned from sophisticated data analysis, enabling them to make well-informed decisions that are in line with changing market conditions. Through the utilization of R's extensive analytical powers, interested parties can predict patterns, enhance their approaches, and guarantee continued participation in the cutthroat world of music streaming.

2. Data Visualization and Analysis

2.1 Figure 1





Visualization Using R

The "Top 20 Best Songs of 2023 by Streams" and "Top 20 Best Songs of All Time by Streams" graphics were made with R, a potent data analysis and visualization program. I created these bar charts by using the ggplot2 tool, which is a popular approach for declaratively constructing visuals. Reading the streaming data from a dataset—which includes details like song titles, artist names, and stream counts—was one step in the process. By aligning the song titles on the x- and stream counts on the y-axes and using the ggplot function to set up the data visually, I was able to improve the graph's readability and visual attractiveness. Next, I used geom_bar to generate bars, each differently colored to represent a distinct artist.

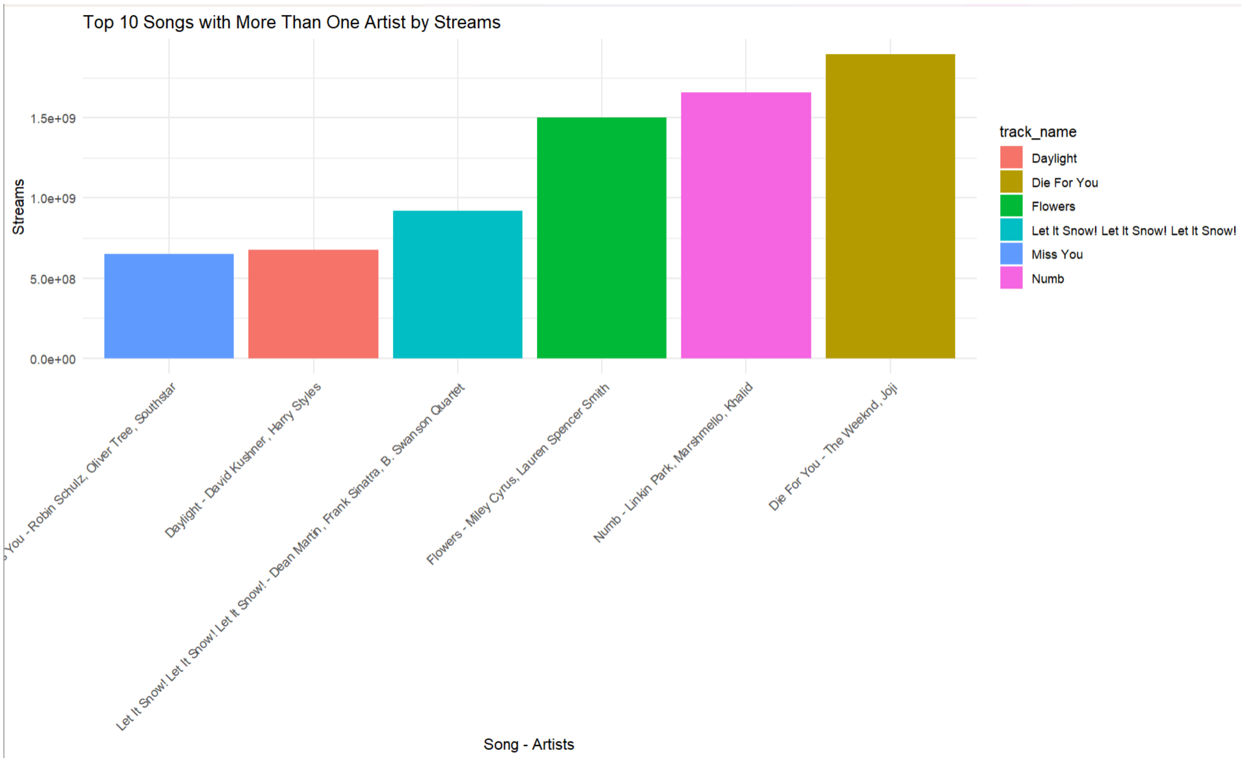
Graph Analysis

These graphics offer a crystal-clear picture of the streaming environment. The "Top 20 Best Songs of 2023 by Streams" graph emphasizes more recent tastes, showing which musicians and genres are currently sweeping the industry, while the "Top 20 Best Songs of All Time by Streams" graph displays a wide variety of artists and genres, representing long-term trends in music popularity. The height of the bars in both graphs indicates the total streams, providing an easy-to-understand comparison of song popularity. With this arrangement, viewers may rapidly determine which songs, both in the past and in the present year, have had the biggest influence on listeners.

Predictive Discussion

By examining these graphs, one can gain an understanding of potential future developments in the music streaming sector. Songs with enduring appeal typically possess timeless elements like catchy melodies or relatable themes, as evidenced by the longevity of select artists and songs on the "All Time" chart. However, the "2023" chart demonstrates abrupt changes in listener tastes, which could be brought about by new music technologies or social media fads. We may anticipate that songs that successfully capitalize on social media's viral potential while preserving a high caliber of music will probably rule the streaming charts in the future. Furthermore, the research indicates that musical preferences are still becoming more diverse, which may inspire producers and musicians to experiment and even work together across genres in order to reach a wider audience.

2.2 Figure 2



Visualization Using R

Using R and the potent visualization package ggplot2, the "Top 10 Songs with More Than One Artist by Streams" graph was produced. The dataset, which is probably kept in a CSV or a related format, provides information on the song titles, artists that are involved, and the total number of streams each song has received. The ggplot function was used to map song titles to the x-axis and stream counts to the y-axis after the data was loaded into R using the read.csv function. For simple visual differentiation between the songs, the bars were colored differently for each one. This method highlights the streaming performance of collaborative tracks by presenting the data in an effective and visually appealing manner.

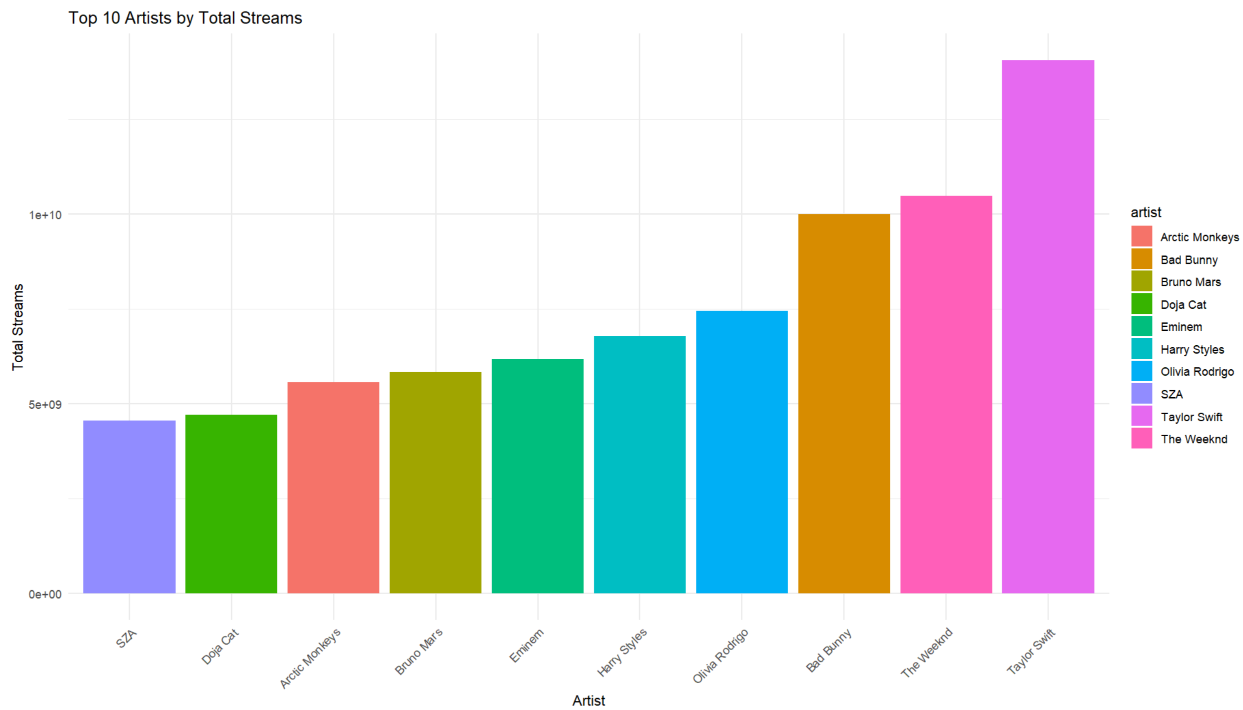
Graph Analysis

The streaming numbers for the top 10 songs with artist collaborations are shown graphically in this bar graph. A distinct song is represented by each bar, and the height of each bar shows the total number of streams. The use of color coding helps to visually differentiate each tune. The graph illustrates the effectiveness of artist collaborations in drawing sizable audiences and is helpful in rapidly determining which joint tracks have had the greatest success on streaming services.

Predictive Discussion

The graph's graphic data highlights how effective artist partnerships are at maximizing song impact and reach. This pattern predictably indicates that artist collaborations, particularly amongst individuals with possibly disparate musical histories or genres, can be an effective way to boost a song's popularity and amount of streams. We may anticipate that record companies and artists will continue to spend in collaborative projects, if not increase it, given the clear success of these partnerships. Future marketing plans may also change to encourage more of these kinds of partnerships, which could result in more cross-genre alliances that appeal to wider audiences.

2.3 Figure 3



Visualization Using R

The "Top 10 Artists by Total Streams" visualization was created with R, a flexible language for statistical computing and visuals. I loaded the data from a pre-compiled CSV file that contains artist names and the accompanying stream counts using the ggplot2 tool, which is well-known for its ability to create sophisticated and visually appealing representations. This data was easier to load into R thanks to the read.csv function. I utilized the ggplot function in conjunction with geom_bar to create this bar chart following the data preparation process, which required ranking and choosing the top ten artists according to stream counts. A unique color was added to each bar to visually distinguish the artists, improving readability and aesthetic appeal. The height of each bar indicates the total number of streams for each artist.

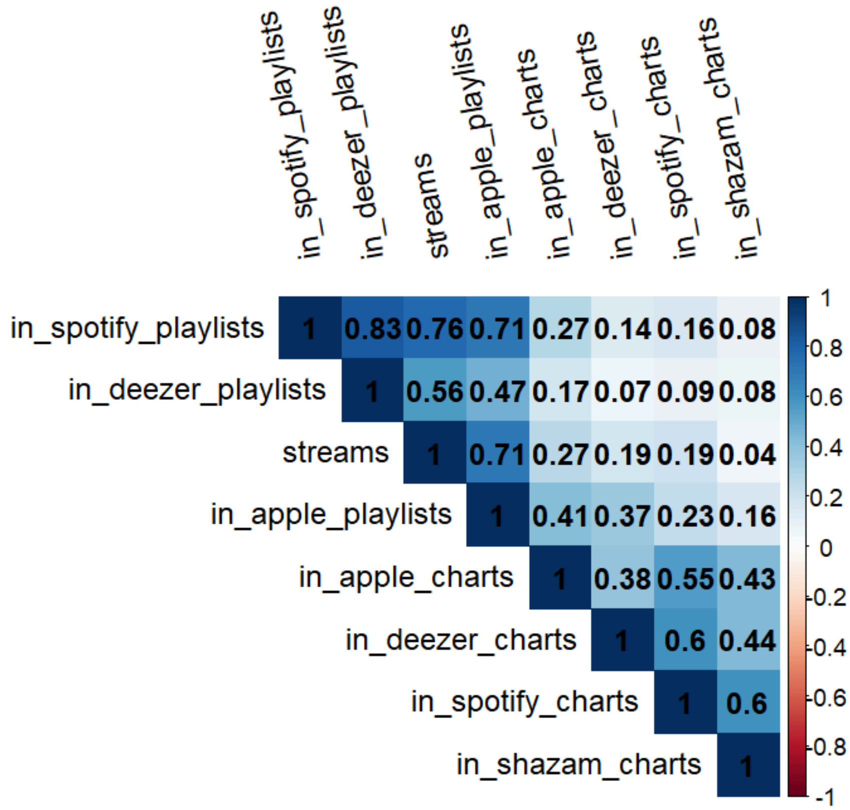
Graph Analysis

The graph efficiently shows the total number of streams for the top ten musicians. Each artist is represented by a different colored bar, making comparisons easier to see right away. The Weeknd and Taylor Swift, who are represented by the highest bars, are the two artists that have the most streaming numbers, which indicates how well-liked and accepted they are in the music streaming market. Based on streaming statistics, this visualization assists in determining which musicians now have the greatest market share and impact in the music industry.

Predictive Discussion

Based on the presented data, it is possible to forecast the continued success of musicians like Taylor Swift and The Weeknd who are regular top performers on streaming charts. Their high streaming numbers indicate a loyal and devoted fan base that will probably continue to support their releases in the future, in addition to their current popularity. This pattern indicates that financial investments in these artists are probably going to pay off handsomely, whether they are made through marketing, partnerships, or additional output. In addition, the range of musical genres represented by these well-known performers suggests a large market where many genres can flourish, indicating a continuous blending of musical inspirations and creativity in the business as musicians fight for the attention of fans in a cutthroat streaming landscape.

2.4 Figure 4



Visualization Using R

R was used to generate the "Correlation between Different Streaming Metrics" correlation matrix, making use of its statistical and graphical features. First, the method involved calculating the correlation coefficients between various streaming measures, like appearance on Spotify playlists, Deezer playlists, Apple charts, and others. The corrplot tool in R is specifically made for visualizing correlation matrices. Initially, the dplyr package was used for data manipulation and the read.csv function was used to load and prepare the data. The Pearson correlation coefficients were calculated using the cor() function after making sure that all pertinent variables had the proper numeric formatting. The corrplot() function from the corrplot package was then used to show these coefficients, making it easier to grasp the results visually by providing a color-coded depiction of the correlation values.

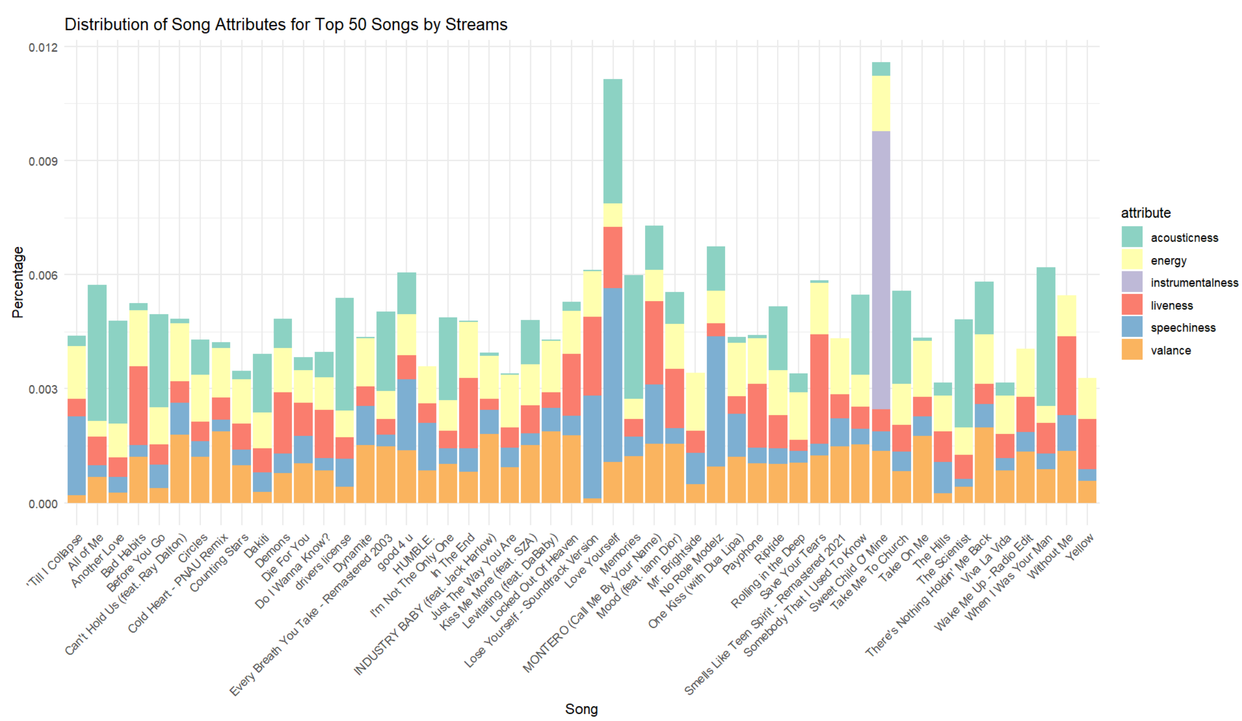
Analysis of the Graph

The correlations between a number of performance-related parameters for music streaming are visually shown by this correlation matrix. Every cell in the matrix shows the correlation coefficient between two metrics, with the strength and direction of the correlations indicated by the colors and sizes of the squares. High positive correlations, for example, are displayed in blue, whereas negative correlations are displayed in red. Strong correlations between related indicators, such inclusion in Spotify and Deezer playlists, are highlighted in the graph, indicating that songs that do well on one platform are probably going to do well on others. Although there is some link between these indicators, other factors obviously have a considerable impact on chart success, as evidenced by the very moderate correlations between playlist inclusions and chart places.

Predictive Discussion

Predictive models in the music industry can be greatly informed by the information obtained from the correlation matrix. The significant connections shown amongst various playlist entries, for example, imply that predictive models may be created based on an artist's or song's performance on a single platform to anticipate success across several platforms. However, the different degrees of correlation between the presence of playlists and chart performance suggest that other factors, perhaps connected to marketing campaigns or listener demographics, also play important roles, indicating that while playlist success is a sign of wider appeal, it is not the only predictor of chart success. By extending the investigation of these aspects, predictive analytics might furnish music creators and marketers with potent instruments to enhance song releases and promotional tactics, ultimately optimizing streaming triumphs and chart rankings.

2.5 Figure 5



Visualization Using R

The "Distribution of Song Attributes for Top 50 Songs by Streams" stacked bar chart was made with R and the ggplot2 package, which is renowned for its robust and adaptable visualization features. In order to prepare the data, pertinent song attributes from a dataset were chosen, including acousticness, energy, instrumentalness, liveness, speechiness, and valence. To enable direct attribute comparisons, each attribute was normalized to reflect its percentage in each song, making sure the numbers added up to 1. 'song' labels the x-axis, 'proportion' scales the y-axis, and 'attribute' chooses the fill color. I use the ggplot function to specify aesthetics by setting aes(x = song, y = proportion, fill = attribute). The properties for each song were stacked on top of one another using the geom_bar(stat = "identity", position = "stack") function, giving a thorough visual overview of each track's composition.

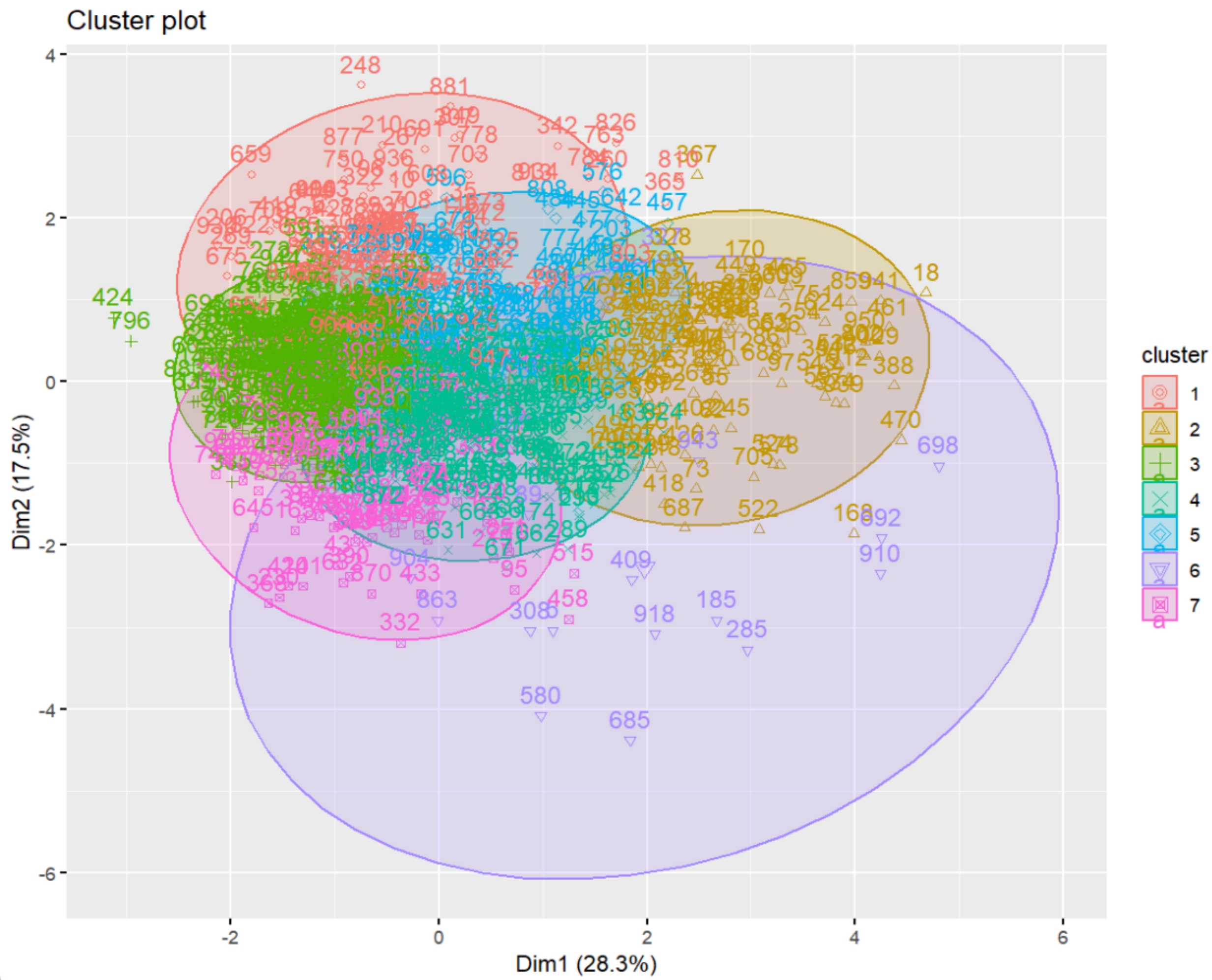
Analysis of the Graph

The graph provides a clear visual representation of how different musical characteristics are distributed across the top 50 songs based on streaming. A single song is represented by each bar, which is divided into colored segments that show the relative amounts of each attribute, such as valence and energy. With the help of this graphic, listeners may instantly see how several famous music differ in terms of liveliness or acousticness. The wide range of musical compositions that appeal to listeners is reflected in the attribute distribution, which ranges from energetic songs to those with prominent instrumental or acoustic aspects.

Predictive Discussion

Predictive insights regarding the potential popularity of a song can be obtained by examining the distribution of song qualities among the top-streamed songs. For example, if a song regularly appears among the top-streamed songs with high valence (positivity) and energy, it could be assumed that future songs with similar qualities will succeed on streaming services. Additionally, record producers and artists can better grasp current musical trends and adjust their works to reflect qualities that are currently appealing to the public with the help of this type of data visualization. Monitoring these characteristics over time may also indicate evolving listener preferences, which could direct the development of musical genres and advertising tactics to foresee and satisfy future consumer needs.

2.6 Figure 6



Visualization Using R

Utilizing a combination of dimensionality reduction and clustering techniques, R was used to construct this cluster plot, which allowed for the analysis of song data according to its auditory properties. The dataset was first processed to add pertinent audio characteristics like acousticness, valence, energy, and danceability. After then, the data was scaled to standardize the range of variables, making sure that scale differences wouldn't allow any one feature to control the clustering process. The clustering was done using the kmeans tool, which provided a number of groups to maximize the grouping based on innate data patterns. Principal component analysis (PCA) was used after clustering to lower the data's dimensionality and enable a 2D representation that effectively conveys the core of the multi-dimensional data. ggplot2 was used to plot the resulting principal components, resulting in this scatter plot. Ellipses encapsulate the general distribution of clusters, and each point represents a song, color-coded according to its partition.

Analysis of the Graph

The first two main components, which account for a sizable amount of the variance in the dataset, are shown against the visual separation of songs into seven unique clusters, each represented by a different hue. Songs with similar qualities are shown by points close together inside each cluster, whilst songs with different qualities are indicated by points far apart between clusters. Cluster cohesion and separation are intuitively conveyed by the ellipses surrounding each cluster, which emphasize their density and spread. Notably, certain clusters are more widely distributed, implying a wider range of qualities inside the cluster, while others are more closely clustered, showing great similarity among those songs. 'Dim1' and 'Dim2' axis are linear combinations of the original features that maximize variance, and they stand for the first and second main components, respectively.

Predictive Discussion

An essential technique for forecasting in the music business is this clustering study. One can forecast the future popularity of new or lesser-known songs based on where they fall within clusters of songs with similar audio profiles. For instance, a new song that belongs to a particular cluster may also be anticipated to do well if songs in that cluster typically have high stream counts. Additionally, by identifying the listener demographics that interact with each cluster's song types most frequently, the clustering can help guide focused marketing initiatives. As additional data becomes accessible, the clustering model can be improved over time and used to update suggestions, target promotional efforts, and even direct the creation of new music in order to conform to the tastes of listeners represented in clusters that are successful.

Part B- Research

3. 1 Four data analytics processes for data governance.

In order to manage the accessibility, usability, integrity, and security of the data in companies, data governance is essential. There are four primary areas into which the data analytics procedures related to data governance may be roughly classified:

1. Management of Data Quality
This guarantees the accuracy, completeness, and dependability of the data utilized throughout an organization. It consists of the following:

- Assessment: Consistently assessing data for dependability, accuracy, and completeness (Mullins, 2017).
- Cleansing: Process of fixing or eliminating unnecessary, inaccurate, or insufficient data.
- Monitoring: Constantly keeping an eye on the quality of data to identify and address problems before they have an impact on company choices.

- Improvement: Creating plans to enhance data gathering, storing, and utilization based on analytics findings (Fisher et al., 2009).

2. Management of Metadata
This process helps data easier to manage and use by assisting companies in understanding the context and structure of their data:
- Collection: Compiling information about the format, meaning, and source of data components.
- Standardization: Keeping metadata uniform and standardized throughout the company.
- Accessibility: Encouraging ethical data management and enhancing data usability by making metadata easily accessible.
- Integration: According to Bradley (2011), metadata should be integrated with business processes to improve data quality and compliance.

3. Management of Data Security
This procedure guarantees that pertinent regulations are followed and safeguards data from unwanted access:
- Policy Development: Formulating thorough security guidelines for data.
- Access Control: Putting in place mechanisms to restrict access to data according to roles.
- Risk management: involves recognizing and evaluating security threats to data.
- Monitoring compliance: Making sure data processes adhere to laws such as HIPAA (Sen & Borle, 2015).

4. Lifecycle Management of Data
Overseeing the data flow from generation to disposal throughout its lifecycle:
- Establishing guidelines and procedures: generate and acquire data.
- Usage: Keeping an eye on how information is utilized within the company.
- Storage and upkeep: Choosing venues for data storage and maintenance procedures.
- Disposal: The safe removal of data that is no longer needed or that is mandated by law to be kept on file (Gartner, 2020).

1.2 What R Has to Offer for Governance Data Visualization

A number of features in R enable data visualization in accordance with data governance principles:
1. ggplot2 for Visualizations of Data Quality: makes intricate graphs to evaluate problems with data quality graphically. Wickham, H. (2016) is cited. The Elegant Graphics for Data Analysis Tool, ggplot2. New York: Springer-Verlag.
2. dplyr for Data Manipulation: Makes it easier to prepare and clean data, which is essential to preserving high data quality. Wickham et al. (2021) is cited. R package 1.0.5, dplyr: A Grammar of Data Manipulation.

3. Plotly for Interactive Dashboards: Facilitates the creation of interactive visual aids that track user behavior and data security. Citation: C. Sievert (2020). R, plotly, and shiny: An interactive web-based data visualization tool. CRC Publishing.
4. data.table for Handling Big Datasets: Helps with effective lifecycle management by supporting the efficient management of big volumes of data. Dowle, M., and Srinivasan, A. (2021) are cited. data.table: CRAN.R-project.org/package=data.table An extension of data.frame
5. dazzling dashboards for real-time monitoring permits the development of dashboards for real-time monitoring, which is helpful for managing data security and compliance. Chang, W. et al. (2021) is cited. shiny: R package 1.6.0, Web Application Framework for R.

2. Ethics is an important aspect of Data visualization

To guarantee that information is presented truthfully, openly, and responsibly, ethics in data visualization is an essential factor to take into account. In addition to promoting credibility and trust, ethical data visualization techniques also guard against disinformation, which can have serious negative effects on the real world. While developing and sharing data visualizations, keep the following important ethical standards and guidelines in mind:

1. Accuracy

Related Practice:
Prior to producing visualizations, always confirm data sources and methods.

Justification:
By guaranteeing the accuracy and dependability of data and techniques, this approach directly upholds the moral precept that honest and correct information should be presented in visualizations (Kirk, 2016). It keeps false or deceptive information from spreading, which is essential to preserving the reliability of the data that is provided.

2. Transparency

Related Practice:
To preserve transparency, provide thorough annotations and information in visualizations.

Justification:
Being transparent entails disclosing the data's origin, the techniques that were used to handle it, and any modifications that were made. Because of this transparency, consumers are able to comprehend the context and constraints of the data, which promotes increased trust and facilitates well-informed decision-making (Few, 2009).

3. Integrity

Related Practice:

Constantly checking my works for any biases or deceptive content.

Justification:
In data visualization, integrity refers to the truthful and objective presentation of data. This procedure preserves the ethical precept of integrity by proactively detecting and correcting any biases or deceptive components, guaranteeing that visualizations are impartial and fair (Cairo, 2015).

4. Accessibility

Related Practice:
Ensuring that all visualizations undergo inclusivity and accessibility testing.

Justification:
Ethical data communication requires that data visualizations be accessible to all users, including those with disabilities. According to Lazar, Feng, and Hochheiser (2017), this approach respects the variety of the audience and encourages inclusivity, making the data accessible to and beneficial to everybody in an equitable manner.

3. Justification on the Part A Prediction

3.3.1 Figure 1

Digital Platforms:
With streaming becoming so popular, it's imperative to keep using these channels to promote music.

Collaborations:
A lot of the top songs are the result of collaborations between musicians, indicating that these kinds of projects are likely to grow in popularity and are an effective way to reach a larger audience.

Decisions Driven by Data:
Because trends in these charts might forecast future musical tastes, the utilization of streaming data can inform decisions about music production and artist promotion.

Marketing Techniques:
Good marketing is still essential since large numbers of streams are frequently associated with successful campaigns.

3.3.2 Figure 2

Strategic Partnerships:
Given the popularity of these joint tracks, record companies and music producers should think about promoting more cross-genre collaborations in order to reach a wider audience.

Marketing & Promotion:
To optimize reach and engagement, collaborative tracks should be supported by focused marketing efforts that take advantage of each artist's distinct appeal.

3.3.3 Figure 3

Artist Promotions:
Record labels may provide priority to new releases, marketing campaigns, and tour support for these highly streamed musicians in order to optimize revenue, given their success with streaming.

Engagement on Streaming Platforms:
By using this data, streaming platforms can create customized playlists and promotions with well-known musicians, which will increase user engagement.

Industry Trends:
Music industry executives can discern patterns in music consumption and artist longevity by observing which musicians are regularly featured on these lists.

3.3.4 Figure 4

Strong Cross-Platform Playlist participation:
Songs that do well on one platform are likely to do well on others, as indicated by the strong correlation (0.83) between a song's participation in Spotify playlists and Deezer playlists. Similar audience preferences and song popularity across platforms can be blamed for this.

Moderate Impact on Streaming Numbers:
There is a noticeable but not overpowering influence, according to the correlations between the presence of playlists and streams (0.71 for Spotify playlists and 0.56 for Deezer playlists). This implies that while a song's streaming numbers are undoubtedly boosted by being on these playlists, other important elements like song quality, artist popularity, and promotional efforts also play a significant influence.

Chart Presence and Streaming:
There are moderate to substantial relationships between chart presence and streaming, especially in the cases of Apple charts (0.37) and Deezer charts (0.55). This suggests that while chart positions reflect streaming success, they are also impacted by extra marketing initiatives and audience interaction.

Platform-Specific Strategies:
To optimize a song's performance and reach, labels and artists should develop customized strategies for each platform, according to the varying levels of correlation observed across them.

3.3.5 Figure 5

Preference for Upbeat and Positive Music:
The graph's high energy and valence prominence indicates that listeners favor energetic and positive music, which suggests that songs with these qualities will probably do well in the future.

Diverse Musical Styles:
The top songs' differences in acousticness and instrumentality reveal a wide listener base with a range of musical preferences. This variety shows that there is still a demand for songs that are both highly produced and acoustic.

Effect on Music Production:
Producers and musicians may choose to concentrate on these elements when writing new songs as a result of realizing that some qualities are associated with higher streams. For example, boosting energy and making sure you're in a good mood could be deliberate goals.

Strategic Marketing and Playlist Placement:
By concentrating on songs that have qualities that match established listener preferences, streaming services and marketers may be able to maximize engagement through playlist placement and promotional activities.

All things considered, this study can help those involved in the music business make well-informed choices regarding songwriting, promotion, and platform interaction in order to take advantage of current trends and improve the success of music streaming.

3.3.6 Figure 6

Basis for Predictive Insights

1. **Cluster Characteristics**:

   A cluster is a collection of songs that have similar aural characteristics. Songs inside a cluster have more in common with one another than with songs in other clusters, as indicated by the tightness and separation of the clusters. The homogeneity seen within the clusters implies that there may be a strong correlation between listener preferences and certain clusters. For example, it's reasonable to assume that other songs in the same cluster will perform well if those songs routinely receive a lot of streaming activity.

2. **Principal Component Analysis (PCA)**:

The most significant variance among the features is retained while the dataset's dimensionality is reduced with the aid of PCA. Simplifying yet effectively expressing song qualities are the first two major components, which account for a significant portion of the variance. Songs' distribution along these plot points can be utilized to uncover underlying patterns in the multi-dimensional space of raw audio data that are not immediately apparent.

3. **Overlap and Distinctiveness of Clusters**:

   The degree to which clusters overlap can reveal whether a music has the ability to appeal to a variety of audience demographics. various and discrete clusters indicate various market segments that are well-defined by particular auditory characteristics. On the other hand, clusters that are adjacent or overlap may have similar listener bases, which may present chances for cross-promotion and multi-cluster targeting in marketing plans.

Predictive Applications

1. **Targeted Marketing and Promotion:**

   Marketers can more successfully adjust their promotional efforts by knowing which clusters are linked to greater popularity or certain demographic preferences. More active promotion of songs from high-performing clusters is recommended, particularly in media and areas where the clusters have demonstrated past success.

2. **New Song Development:**

   When creating new music, producers and musicians might draw inspiration from cluster qualities. If some characteristics appear frequently in clusters that are successful, adding these characteristics to new tracks may make them more likely to succeed.

3. **Recommendation Systems**:

   By adding cluster-based insights, music streaming services can improve their recommendation systems. The system can suggest more songs from the same cluster to users based on their preferences for particular clusters, which increases user pleasure and engagement.

4. **Future Trend Analysis**:

   Tracking songs as they migrate across clusters over time or as new clusters form can give early warning signs of changing musical trends. For industry participants to remain ahead of changes in the market and modify their strategy appropriately, this can be quite important.

To summarize, cluster analysis is a useful technique for comprehending present-day song attributes and preferences. It also yields a prediction model that can foresee future trends and behaviors in the music streaming industry. The music industry can now formulate strategies in a proactive manner thanks to this predictive capability, which is based on data-driven insights.

4. Conclusion

Significant patterns that might inform forecasts and business plans in the music industry are revealed by the examination of several datasets pertaining to music, such as the top songs by streams, artist collaborations, and song qualities. Data on the most popular songs and artists on streaming services makes it clear that a wide range of musical genres and well-thought-out partnerships are key factors in streaming success. The Weeknd and Taylor Swift are two artists that rule the music scene by fusing a broad appeal with regular, excellent releases.

By demonstrating how particular attributes, such as energy, valence, and appearance on playlists across several platforms, interact with streaming performance, the correlation matrix and attribute distributions significantly advance our understanding. Songs with a high level of energy and emotional positivity, or high valence, typically receive more streams. Furthermore, the fact that an artist is featured on popular playlists and charts on platforms such as Spotify, Deezer, and Apple Music highlights the significance of incorporating a digital strategy into their promotional endeavors.

Artists and record labels are expected to depend more and more on data-driven insights in the future for decision-making. Expect a persistent focus on collaborating with other artists to broaden their reach and producing music that adheres to the qualities of high streams. Furthermore, it will continue to be essential to leverage digital media effectively for focused music marketing. Future trends in music creation and marketing will probably be determined by this all-encompassing strategy that is based on analytics.

Reference List

1. Bradley, N. (2011). *The Management of Metadata.* In: J. G. Williams and S. L. Hansen (Eds.), Information Management for Business and Competitive Intelligence and Excellence. Springer, pp. 23-36.
2. Cairo, A. (2015). *The Truthful Art: Data, Charts, and Maps for Communication.* New Riders.
3. Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2021). *shiny: Web Application Framework for R.* R package version 1.6.0. Available at: https://CRAN.R-project.org/package=shiny
4. Dowle, M., & Srinivasan, A. (2021). *data.table: Extension of data.frame.* Available at: https://CRAN.R-project.org/package=data.table
5. Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis.* Analytics Press.
6. Fisher, T., Lauria, E., Chengalur-Smith, S., & Wang, R. (2009). Introduction to Information Quality. MIT Information Quality Industry Symposium, pp. 13-20.
7. Gartner. (2020). *Data Lifecycle Management: The Need for Comprehensive Data Management Strategies.* Available at: https://www.gartner.com/en/documents/3989463
8. Kirk, A. (2016). *Data Visualisation: A Handbook for Data Driven Design.* Sage.
9. Lazar, J., Feng, J.H., & Hochheiser, H. (2017). *Research Methods in Human-Computer Interaction.* Second Edition, Morgan Kaufmann.
10. Mullins, C. S. (2017). *Database Administration: The Complete Guide to Practices and Procedures.* 2nd ed. Addison-Wesley Professional.
11. Sen, R., & Borle, S. (2015). *Data Protection and Privacy in the Age of Big Data.* Statistical Analysis and Data Mining: The ASA Data Science Journal, 8(4), 245-259.
12. Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny.* CRC Press.
13. Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York.