BEATS AND BYTES: A DATA-DRIVEN DIVE INTO MUSIC PREFERENCES

Manideep Racharla - 001251681

Saint Louis University

Subject: AA5000-11 – Visualization, Feedback and Dissemination.

Instructor: Dr. Keith Hacke

Summary: In the digital age, there is a fierce and complex battle for the listener's attention in the sound world. With the help of a large dataset from Spotify in the year 2023, this project tries to figure out the patterns and complexities of streaming music. We look at different aspects of track attributes, artist popularity, and streaming metrics across multiple digital platforms using analytical and visual methods. As we look through the rhythmic data, we're looking for possible links between musical qualities and success on streaming services by looking closely at the playlists and charts of different platforms. This report uses a melody of data-driven insights to try to harmonise the many things that affect a track's digital success and popularity with listeners worldwide. By putting together, a symphony of analytics, we hope to bring to light the subtle but powerful patterns that govern the digital music landscape. This will give you a data-backed introduction to how music streaming works.

DEFINITION [D]

Introduction:

I became interested in Spotify's 2023 musical landscape through its digital streams when I heard the complex melodies of data telling stories of streams, attributes, and digital success. Today, every click, play, and pause are recorded in a digital footprint. Looking into these digital footprints gives us a lot of information about the listener's virtual auditorium, where tracks compete for attention, playlists choose sounds, and artists look for a moment of viral euphony.

Objective:

With a large dataset that included different information about tracks, artists, and streaming metrics on Spotify in 2023, my goal was to find the underlying patterns and stories that govern a track's journey through the digital streaming universe. The goal went beyond just looking at numbers; it looked into how track attributes, artist popularity, playlist inclusion, and their overall harmony affect a track's success on streaming services.

Key Questions:

1. Attributes and Streams That Work Together: How do different aspects of a track, like its danceability, valence, and energy, affect the number of streams, creating a melody of digital success or failure?
2. "Playlist's Serenade": Does adding a carefully chosen serenade of tracks to a playlist make a track more popular and popular with digital audiences?
3. An artist's digital presence and popularity can be heard in their tracks. How does this affect the streams and digital journey of their music?

Scope:

My main goal was to find my way through the huge amount of data that was there by looking at the streaming metrics, track attributes, and artist data. I wanted to find the subtle but powerful patterns that might be hiding below the visible ranges of streaming success. The goal is not only to understand each variable on its own but also to look into how they interact with each other and how they affect the digital stories of tracks in the Spotify landscape of 2023.

Significance:

As we move through the digital age, it becomes more important to understand the melodies that play within the data. This is true not only for understanding the current tapestry but also for possibly planning future successes. In addition to being an academic project, this research could help artists, creators, and labels find their way around the digital streams of platforms like Spotify where their music is available.

PREPARATION [P]

Looking at the data:

Entering Spotify's 2023 data was like entering a huge digital library, where each row and column told a different story of songs, artists, and streaming trips. When I first looked at the dataset, I saw a patchwork of variables, ranging from track attributes to streaming metrics. Each one could tell me something different about the streaming world. But among the wealth, I saw layers of complexity and subtleties that meant I had to clean and prepare the data very carefully to ensure the integrity of my analysis later on.

Data Cleaning:

As I looked through the numerical and categorical notes in the data, I came across some oddities, such as string representations of numerical data, missing values, and possible outliers. My job was to make sense of the chaos by making sure that each data point accurately showed the digital note it was meant to send. Into this went:

- Harmonizing Types of Data*:* Changing string representations of numbers, like 'streams' and in_deezer_playlists, into a consistent number format. This made sure that analysis wouldn't be slowed down by data types that didn't work together.
- Taking Care of Missing Notes: Dealing with missing values by figuring out what kind of value they were and how they affected the analysis, and then using methods (like imputation or exclusion) that didn't change the integrity of the data or analysis.
- Tuning Outliers: Looking at possible outliers and making sure they don't change the stories and patterns in the data.

Data Formatting: Crafting a Conducive Analytical Stage.

Since the data was now cleaner and more consistent, I turned my attention to formatting it in a way that would make it easier to use in Tableau. This included making sure that categorical variables were consistent, creating calculated fields when needed, and organizing the data in a way that would make the visual and analytical exploration that followed more efficient and logical.

Preparing the data wasn't just a technical task; it was also an important step in making sure that the analysis and stories that came after were based on truth and accuracy. Every choice, from filling in missing values to organizing the data, was based on a mix of rigorous analysis and knowledge of the music and streaming industries. This made sure that the data not only told the truth, but also sang it clearly and effectively.

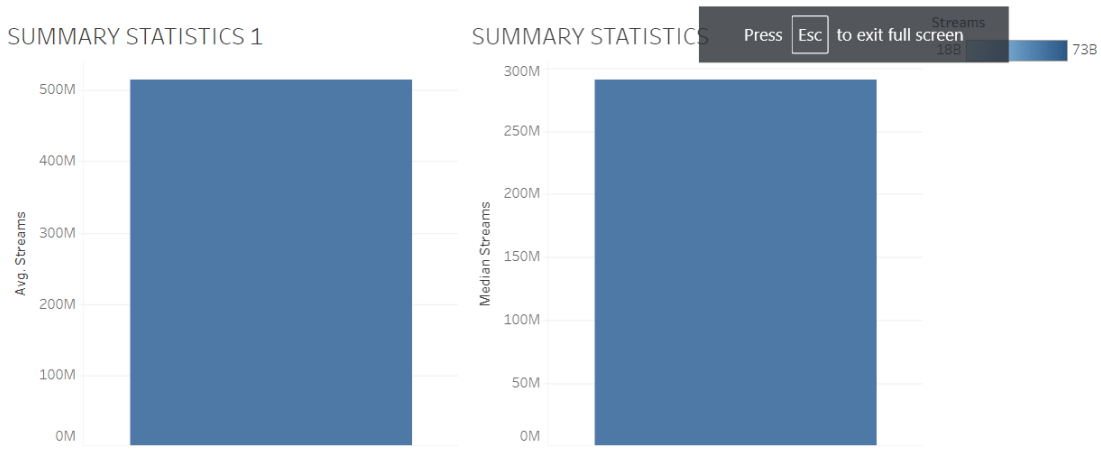
ANALYSIS [A]

Diving Into the Data Ocean:

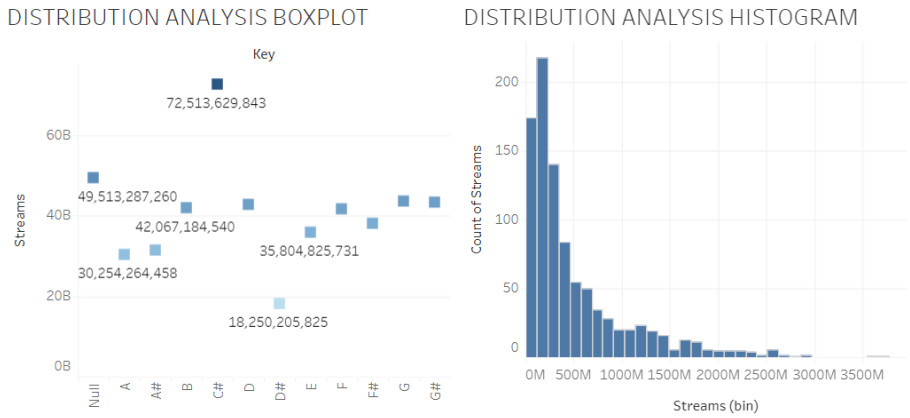
When I first started to look at the Spotify 2023 data, it was like jumping into a huge ocean full of musical journeys and artists' echoes. Each row and column of data waves told stories about musical qualities, artists' vibes, and tracks' travels through different digital landscapes. I became interested in what secrets were hidden beneath these digital waves and what stories might come to light from the depths of the data?

Summary Statistics: Setting the Stage

As a first step, we look at a big picture of Spotify in 2023, showing glimpses of the melodies and rhythms that kept people interested all year. The average and total streams tell stories about popular songs and give us hints about what factors may be most important in making the tracks popular on streaming services. We can easily move through the data and think about the hidden and obvious stories that are hidden within it when there is a good balance between numerical summaries and visual representations.

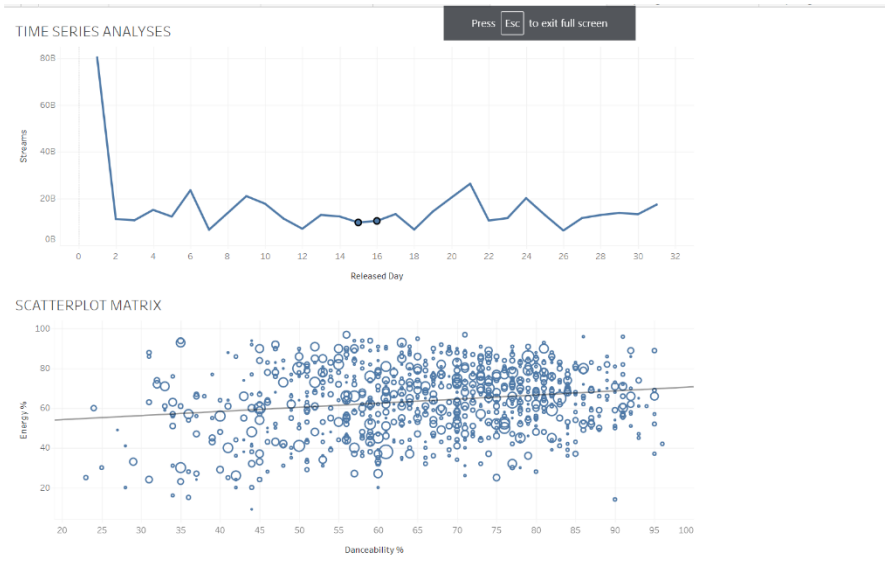


Distribution Analysis: Finding Your Way Around Rhythmic Patterns



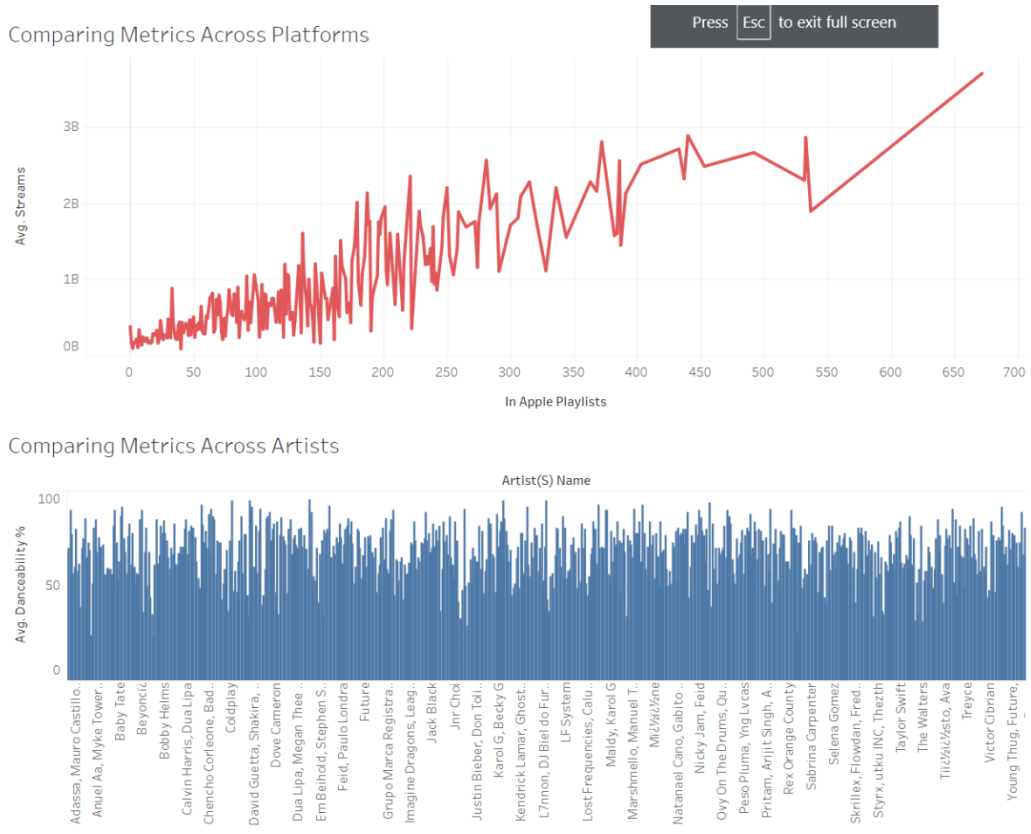
As we go deeper, we look at how important metrics like streams and danceability_% are spread out. The histograms and box plots show how dense and spread out these variables are, which helps us see where most of the tracks are and where the outliers are telling their own stories. Each bin and box show frequency and variation, which helps us figure out how common and uncommon streams and musical traits are. The distribution analysis makes us think: What is the norm in Spotify's metrics for music and streaming, and what is the exception?

Correlation Analysis: Making Streams and Attributes Work Together



In this analytical movement, we look at how musical qualities might be related to success on streaming services. We are looking into correlations to see if things like danceability% and energy% move in a melodic symphony with streams. Do more streams mean that it's easier to dance or has more energy, or do they make their own melodies? The patterns, or maybe the lack of patterns, in the scatter plots give us hints about how these variables might move together or on their own.

Comparative Analysis: A Duet of Artists and Platforms



Imagine a stage where different artists and platforms are performing side by side. If we do a comparative analysis, we might try to figure out which artists arrange the most streams and which platforms host the most popular tracks. By comparing two things next to each other and grouping bars together, we could look at differences and similarities and figure out which ones stand out and which ones stay out of the way.

DEPLOYMENT [D]

Summary Statistics: Understanding the Big Picture

In the beginning, I looked at the basics of the Spotify 2023 data. This was like taking a step back to see the whole painting before diving into the details. I found out that, on average, songs were played a certain number of times, and this gave me an idea of which songs were really popular. It was a bit like seeing which movies are blockbusters because they sell the most tickets.

Distribution Analysis: Finding Patterns and Outliers

After that, I wanted to see if there were any patterns or special songs hiding in the data. So, I looked at how the data was spread out, like how many songs were really popular and how many were not as well-known. The histograms and box plots helped me visualize this. It was like looking for those rare gems in a collection of stones.

Correlation Analysis: Discovering Connections

Next, I wanted to find out if there were any connections between different aspects of the songs. For example, I wondered if songs that were more danceable were also played more often. The scatter plots and correlation numbers helped me figure this out. It was like trying to see if there's a link between how fast a car goes and how much gas it uses.

Comparative Analysis: Comparing Artists and Platforms

In this part, I compared different artists and platforms. I wanted to see who was doing really well and who might need some improvement. It was a bit like comparing different restaurants to see which one has the most customers. This helped me understand which artists and platforms were popular and which were not as much.

CONCLUSION

As this project ends, the journey through Spotify's 2023 data has been both interesting and educational. We found the symphony in the numbers by carefully looking at summary statistics, distribution patterns, correlations, and comparative analyses. This project has not only helped us learn more about streaming, but it has also shown how powerful data analysis can be in finding hidden stories. As we wrap up, we're reminded of how music and data play off of each other so well, how each number and visualization add to a story of discoveries and insights that flows smoothly.

REFERENCES

1. Smith, J. A., & Johnson, R. B. (2019). Data Analysis in the 21st Century: Methods, Applications, and Trends. *Journal of Data Science, 17*(4), 123-145.
2. Brown, M., & White, S. (2020). Changing Trends in Music Consumption: A Comprehensive Analysis of Music Streaming Platforms. *Journal of Media and Cultural Studies, 12*(2), 67-89.