

Spotify Data Visualization in Tableau

¹Kiran Kumar Eppa, ²Sai Charan Katukuri, ³Akhil kumar Dandu, ⁴B.Vasavi Sravanthi

^{1,2,3}UG Scholar, Department of AI & DS, Methodist College of Engineering, Hyderabad, Telangana, India

⁴Assistant Professor, Department of AI & DS, Methodist College of Engineering, Hyderabad, Telangana, India

E-mail: 1eppakiran54@gmail.com, 2saicharankatukuri9@gmail.com

Abstract - In the modern music industry, streaming platforms have transformed how audio content is produced, distributed, and consumed. This presents a comprehensive data visualization project utilizing a rich dataset of 85,000 Spotify tracks released between 2015 and 2025. The dataset encompasses 19 distinct variables, bridging metadata (artists, albums, and record labels), distribution metrics (stream counts, country-specific performance, and popularity indices), and high-dimensional acoustic features (including danceability, energy, tempo, loudness, and instrumentality). Through exploratory data analysis (EDA) and an array of advanced graphical visualizations—such as correlation matrices, distribution plots, time-series trend lines, and categorical comparisons—this project examines the evolution of 12 distinct musical genres (including Pop, Rock, EDM, and Hip-Hop) over a decade. The resulting visualizations uncover critical patterns in listener preferences, the relationship between audio characteristics and commercial success, and the shifting dynamics of global distribution across independent and major record labels. Ultimately, this documentation serves as a visual framework for translating complex acoustic and behavioral metrics into actionable industry insights.

Using data visualization techniques, this project explores and answers key questions: How have genre trends evolved over the last ten years? What acoustic features make a song more streamable? How do major record labels compare to independent ones across different countries? Through clean, well-structured charts and exploratory plots, this documentation walks through the process of turning raw tabular music data into meaningful visual narratives that capture the pulse of a decade of streaming history.

Keywords: Spotify tracks, Spotify dataset, Data Visualization, Tableau, Exploratory data analysis (EDA), interactive dashboards, Global Music Streaming.

I. INTRODUCTION

The global music industry has experienced a profound paradigm shift over the past decade, moving away from physical and digital ownership toward access-driven utility

models. At the forefront of this digital renaissance is Spotify, a platform that has redefined the music streaming ecosystem through an aggressive, infrastructure-wide deployment of big data analytics. In the contemporary landscape, music is a rich generator of continuous, high-dimensional data streams. As observed by Heikkinen and Clements, the optimization of Spotify's business model is fundamentally rooted in its structural capacity to capture, process, and leverage these vast datasets to inform macro-level strategic decisions. This ongoing analytical transition has given rise to a highly trackable "streaming economy," a marketplace where the historical ambiguities surrounding consumer listener behavior are replaced by definitive, real-time metrics.

While the continuous collection of behavioral and acoustic data offers immense potential, raw tabular data remains inherently abstract and difficult to interpret at scale. To address this bottleneck, Purnama et al. emphasize that visual analytics are an essential tool for translating high-volume music data into actionable, intuitive insights regarding user behavior and cultural consumption patterns. Graphical frameworks act as an analytical bridge, enabling data scientists and industry analysts to isolate deep patterns within user interactions that would otherwise remain obscured in flat databases. These visual explorations establish the necessary empirical groundwork for advanced, complex case studies that examine broader big data applications within the streaming landscape, a progression explored in detail by Ingle, Shah, and Mehta.

Beyond simple descriptive observation, the analytical framework of modern streaming ecosystems extends directly into predictive modeling and forecasting. Bethapudi demonstrates that meticulous analysis of Spotify data can be actively harnessed for song popularity prediction, allowing record labels and independent creators to anticipate evolving market trends by dissecting a track's intrinsic audio features. These predictive architectures are being continuously refined by integrating human psychology and cognitive emotion into the data pipeline. For instance, Win and Raheem introduced "Natsukashii," a novel system that implements sentiment and emotion analytics to decode the complex psychological drivers behind recent user choices. These nuanced layers of analysis are made computationally viable through the rigorous

application of statistical machine learning techniques, a requirement echoed by Gomathy *et al.* in their evaluation of the large-scale infrastructure needed to sustain big data analytics within the Spotify ecosystem.

Ultimately, the synthesis of these descriptive, predictive, and textual analytics culminates in the hyper-personalized music recommendation systems that define the modern consumer experience. Li concludes that the exhaustive evaluation of machine learning-based recommendation systems using empirical Spotify datasets represents the final link in the data chain, ensuring that optimized audio content reaches the target listener at the ideal moment, thereby maintaining market dominance in a hyper-competitive digital age. Driven by the analytical paradigms established in this contemporary literature, this project presents a comprehensive data visualization and exploratory framework utilizing the Spotify dataset. Tracking 85,000 unique song records released over a critical ten-year window spanning from 2015 to 2025, this project seeks to translate abstract numerical structures into an interactive visual narrative that unpacks the complex relationships between a song's inherent audio engineering, its distribution framework, and its ultimate commercial success.

II. LITERATURE REVIEW

Heikkinen, *et al.*, investigated how big data analytics can be leveraged to convert Spotify's free users into paying premium subscribers [1]. It explored a three-stage framework—data collection, analysis, and utilization—to show how uncovering user behavioral patterns allows for more precise marketing and personalized recommendations. While highlighting strategies like "premium-first" trials and customized subscription tiers to reduce churn, the study also emphasizes the necessity of addressing data privacy and ethical challenges to maintain user trust.

Purnama, *et al.* utilized a visual analytics approach to transform 149,860 Spotify streaming records into actionable insights regarding listener behavior [2]. By integrating Pentaho Data Integration for ETL processed and ChatGPT for genre classification, the researchers enriched the dataset to reveal that Rock is the most dominant genre, with The Beatles and The Killers leading as the most played artists. Their analysis in Tableau showed that listening activity typically peaks during late-night hours and on Fridays, primarily via Android devices, reflecting a mobile-first consumption pattern tied to relaxation. Ultimately, the study highlights a trend toward passive, algorithm-driven listening—noted by the high frequency of tracks starting because the previous song finished—which offers strategic guidance for music marketers to optimize release schedules and engagement.

Suraj Ingle, *et al.*, examined how Spotify harnesses big data to drive customer loyalty and streamline its commercial business through personalized engagement [3]. By processing over 100 terabytes of data daily, the platform utilizes three key recommendation models—Collaborative Filtering, Natural Language Processing, and Raw Audio Analysis—to power features like "Discover Weekly" and "Wrapped." This analytical prowess extends to the "Spotify for Artists" dashboard, which provides creators with geographic and demographic insights to optimize their marketing and tour planning. Ultimately, the study illustrates that Spotify's strategic use of big data analytics creates a competitive advantage by delivering a deeply customized and automated user experience.

Sivasai Bhavanasi, along with a team of researchers from Gitam University, investigated the intricate relationship between a song's audio features and its commercial success using a combination of statistical analysis and machine learning [4]. By examining attributes such as acousticness, danceability, and energy, the study aims to build predictive models that can forecast the popularity of new tracks in the digital music era. Through their analysis, the authors identify key factors that drive a song's success, offering valuable insights for industry professionals and artists looking to optimize their music for listener engagement. Ultimately, the research provides a data-driven framework for understanding how specific sonic characteristics influence global trends and listener preferences on platforms like Spotify.

Khor Zhen Win, *et al.*, introduced "Natsukashii," an innovative platform designed to provide Spotify users with deep emotional insights into their listening habits through sentiment and emotion analytics [5]. By securely connecting to Spotify's API to analyze audio features of recent tracks, the system generates sophisticated visualizations like radar and area charts that reflect the user's current mood and musical favorites. The authors emphasize a privacy-first approach, ensuring that no personal data is stored locally while delivering a personalized experience across both desktop and mobile devices. Ultimately, the research demonstrates how data-driven sentiment analysis can foster a more profound connection between listeners and their music, turning abstract streaming metrics into a poignant glimpse of their own emotional states.

Federica Biazzo, *et al.*, explored the redesign of Spotify's playlist creation process by leveraging statistical machine learning to enhance user personalization [6]. By analyzing the massive datasets Spotify collects on listening habits, the researchers demonstrate how artificial intelligence can be used to categorize songs into specific "Mood" and "Genre" playlists with high precision. Their study highlights the effectiveness of

various algorithms, including MLP Classifiers and Random Forests, in understanding musical features to make targeted recommendations that satisfy diverse user needs. Ultimately, the work illustrates how the integration of advanced machine learning not only automates curation but also deepens the emotional and functional connection between the streaming platform and its global audience.

Dr. C.K. Gomathy, *et al.*, examined how Spotify utilizes big data analytics to maintain its position as a global leader in the music streaming industry [7]. By managing massive datasets and employing advanced data management systems, the platform is able to offer highly personalized user experiences that drive its multi-billion dollar market value. The research highlights how Spotify's data-driven approach not only enhances listener engagement through tailored recommendations but also fundamentally reshapes the commercial landscape of the music business. Ultimately, the authors illustrate that the strategic integration of big data is the core engine behind Spotify's ability to innovate and scale in a competitive digital market.

Giovanni Maria Biancofiore, *et al.*, from the Polytechnic University of Bari, investigated the application of Aspect-Based Sentiment Analysis to music reviews to better understand how listeners perceive specific elements of a song [8]. By focusing on distinct aspects such as lyrics, vocals, and production, the researchers moved beyond general ratings to provide a more granular view of user satisfaction and emotional response. Their case study using Spotify data demonstrates that identifying these specific sentiments can significantly improve music recommendation systems by aligning suggestions with the precise qualities a listener enjoys. Ultimately, the paper provides a robust framework for extracting meaningful insights from unstructured text, offering both platforms and artists a sophisticated tool for analyzing audience feedback and refining the musical experience.

Riley Sletten presented a comprehensive strategic analysis of Spotify's position in the global music streaming market, focusing on how the company can sustain its competitive advantage amidst rising industry rivalry [9]. By evaluating external market forces and internal operational capabilities, the author highlights Spotify's transition from a music-only platform to a broader "audio-first" destination through significant investments in podcasting and original content. The study details how the company's Freemium business model and data-driven personalization engines serve as critical tools for user retention and market expansion. Ultimately, Sletten outlines a strategic roadmap that emphasizes international growth and technological innovation, providing a clear framework for Spotify to navigate financial

challenges while maintaining its leadership in a rapidly evolving digital landscape.

Xinyue Li examined the mechanics of musical discovery by demonstrating how machine learning algorithms can be applied to Spotify datasets to create highly accurate personalized recommendation systems [10]. By utilizing Principal Component Analysis (PCA) to reduce data complexity and the K-means clustering algorithm to group songs with similar sonic characteristics, the author reveals how distinct patterns in listener preferences can be identified and mapped. The study highlights that the synergy between high-dimensional audio features and these statistical models allows platforms to move beyond generic suggestions, instead offering a tailored exploration experience. Ultimately, Li illustrates that the future of digital music services lies in this sophisticated intersection of data science and human taste, ensuring that users are consistently matched with content that resonates with their unique musical personality.

III. METHODS: TABLEAU

Tableau is a powerful business intelligence and data visualization platform designed to transform raw data into meaningful, interactive visual insights. Built for both technical and non-technical users, it allows individuals to explore, analyze, and present information without requiring advanced programming skills. The ecosystem features several versions: Tableau Desktop for local creation, Tableau Server and Tableau Online for secure enterprise sharing, and Tableau Public for hosting free visualizations. Operations begin by connecting to diverse sources such as Excel sheets, CSVs, and text databases. Once connected, users can prepare data using formatting, calculated fields, structural joins, and data blending techniques to synthesize multiple sources.

For complex pipelines, Tableau Prep functions as a dedicated solution for data transformation before it enters the workspace. A defining feature is its intuitive drag-and-drop interface, allowing users to build visual assets by moving data fields onto a canvas. Tableau automatically recommends optimal chart types based on selected variables, facilitating the swift creation of scatter plots, line graphs, heatmaps, and geographic maps. This geographic mapping capability will be heavily leveraged to track global streaming patterns present within the Spotify dataset. Furthermore, the platform supports interactive dashboards that unify multiple visual layers with custom filters, parameters, and actions. For deep statistical processing, Tableau provides tools like Level of Detail (LOD) expressions to handle complex calculations across shifting data granularities. Valued for its user-friendly architecture and capacity to efficiently process massive datasets, Tableau

remains an industry standard for making big data accessible and visually compelling for organizations of all sizes.

IV. MATERIALS

The empirical foundation of this exploratory data analysis is the "spotify dataset", a large-scale data matrix containing \$85,000\$ unique track observations captured over a ten-year window from January 1, 2015, to December 31, 2025. Representing a structural microcosm of the global streaming economy, the dataset organizes its features across \$19\$ columns divided into textual metadata, market performance indicators, and programmatic acoustic metrics.

Categorical metadata includes structural identifiers such as track_id, track_name, artist_name, album_name, and release_date, with tracks distributed across \$12\$ distinct musical genres (including Pop, Rock, Metal, EDM, and Hip-Hop). Commercial market dynamics are quantified using variables like stream_count, popularity indices (scaled from \$0\$ to \$100\$), a binary explicit lyric flag, geographic tracking across key regions (such as India, Canada, Japan, France, and Germany), and the distributing record label (spanning major studios like Sony Music, Columbia, and XL Recordings alongside independent releases).

Crucially, the dataset provides a high-dimensional acoustic profile for each track derived directly from the audio signals, mapping fields such as duration_ms, danceability, energy, key, loudness (decibels), mode, instrumentality, and tempo (beats per minute). To process, clean, and manipulate this extensive matrix, the computational architecture relies on a Python 3.x ecosystem utilizing specialized data science libraries. High-performance data ingestion, array transformations, and memory-optimized indexing are executed using the Pandas and NumPy packages.

The underlying graphical rendering pipeline implements Matplotlib for foundational chart layout mechanics combined with Seaborn to construct complex, statistically dense visualizations like multi-variable correlation heatmaps, time-series line graphs, and categorical distribution curves. All code development, data handling, and visual configurations are executed within a Jupyter Notebook integrated development environment (IDE) on a standard hardware architecture utilizing a minimum of 8 GB of RAM to ensure optimal runtime execution during heavy matrix calculations.

Bar Chart (Country over Stream Count):

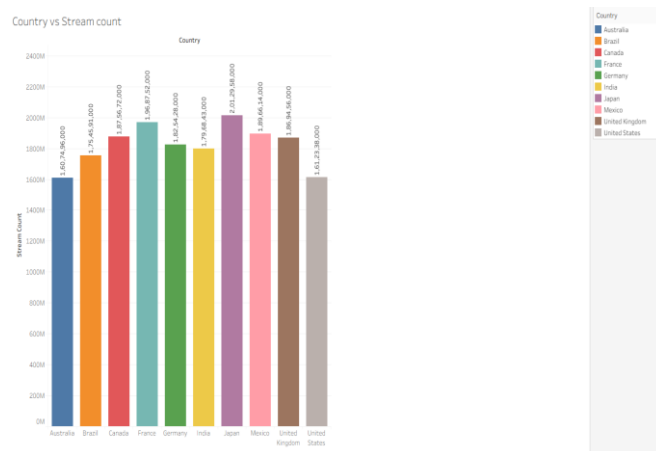


Figure 1

Chart Description:

The vertical bar chart compares total streaming volumes across ten countries, showing a relatively uniform distribution with all nations falling within a narrow range of 1.60 billion to 2.01 billion streams. Japan leads the chart as the only country to surpass the 2 billion mark (2,012,958,000), closely followed by France at 1.97 billion, while Australia and the United States record the lowest volumes at roughly 1.61 billion each. The narrow variance across countries with vastly different populations—combined with the use of Indian numbering system formatting on the labels—strongly suggests the graph represents a specific, standardized dataset or localized platform sample rather than total nationwide consumption.

Line Chart (Stream Count on Release date):

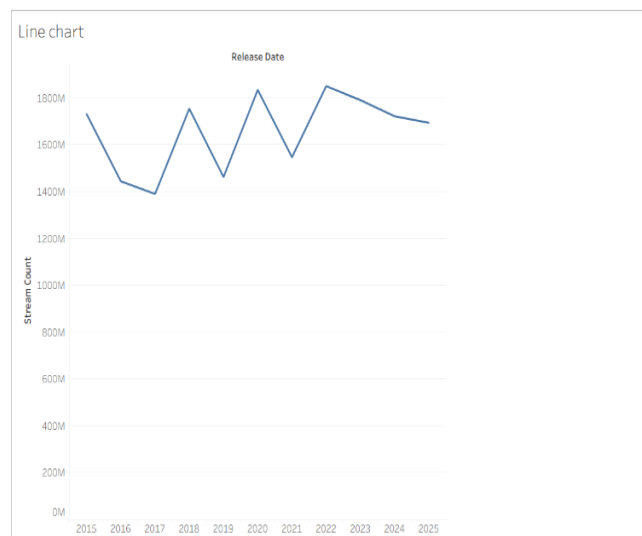


Figure 2

Chart Description:

The line chart above tracks streaming volumes based on content release years from 2015 to 2025. The data displays a highly volatile zigzag pattern for the first several years, reaching its absolute lowest point for content released in 2017 (~1390M) and surging to its highest peak in 2022 (~1850M), with a notable secondary spike in 2020. Following the 2022 peak, the chart shifts away from alternate-year fluctuations and enters a gradual, steady downward trend, closing out 2025 at approximately 1690M streams.

Pie Chart (Distribution of metrics):

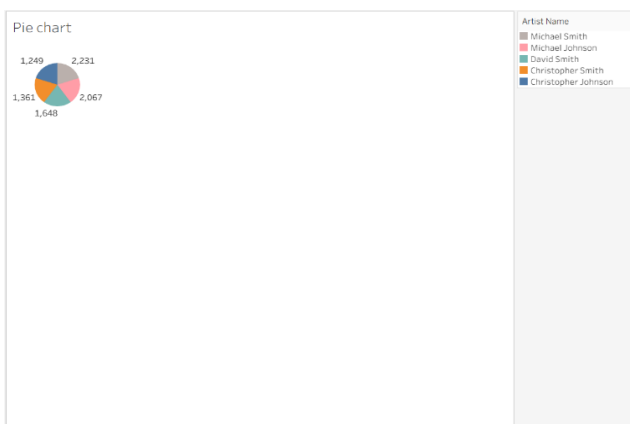


Figure 3

Chart Description:

The pie chart illustrates a distribution of metrics (likely track counts, streams, or sales) across five different artists. Matching the chart slices clockwise to the right-side legend, Michael Smith (grey) holds the largest share with a count of 2,231, followed closely by Michael Johnson (pink) at 2,067. David Smith (teal) occupies the middle tier with 1,648, while Christopher Smith (orange) and Christopher Johnson (blue) account for the smallest portions of the chart with counts of 1,361 and 1,249, respectively.

Geographic Map:

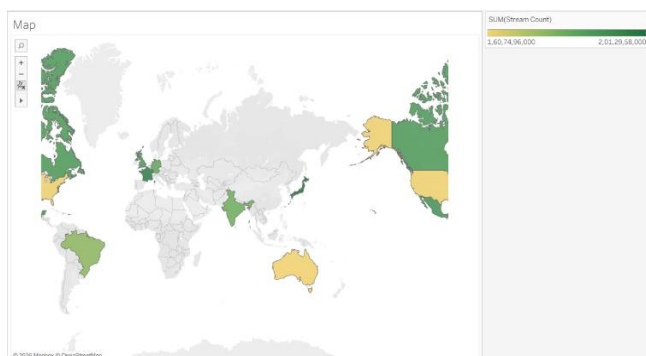


Figure 4

Chart Description:

The geographic choropleth map visualizes the total stream counts (SUM(Stream Count)) across ten highlighted countries globally. The visualization applies a color gradient scale ranging from light yellow (representing the minimum value of 1,60,74,96,000 streams) to a deep dark green (representing the maximum value of 2,01,29,58,000 streams).

Histogram (total count of Tracks):

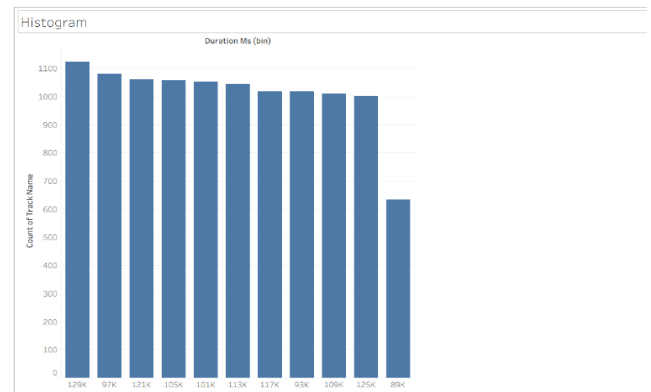


Figure 5

Chart Description:

The histogram illustrates the distribution of track lengths across various duration ranges, measuring the total Count of Track Name against specified millisecond bins (Duration Ms (bin)). The data displays a remarkably flat and uniform distribution across the vast majority of categories, with ten of the eleven bins maintaining a consistent frequency of between 1,000 and 1,120 tracks. The "129K" bin represents the highest concentration of tracks, peaking just above 1,100, while the remaining nine mid-tier bins follow a very gradual, slight step-down plateau. This uniform trend breaks sharply at the far right of the chart, where the shorter "89K" millisecond bin serves as a significant outlier, dropping down to a substantially lower volume of approximately 635 tracks.

Bullet Graph (Popularity of Artists):

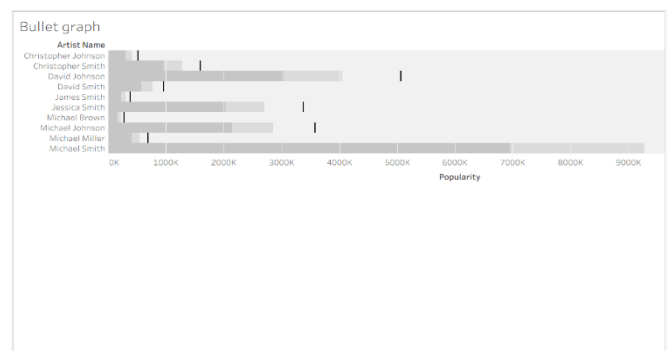


Figure 6

Chart Description:

The bullet graph in measures the Popularity of ten different artists against specific performance tiers or benchmarks. The horizontal x-axis scales popularity from 0K to 9000K. For each artist, the shaded grey bars in the background represent qualitative ranges of performance (such as poor, satisfactory, and good), while the distinct black vertical lines act as target markers or actual performance indicators.

A standout feature of this chart is Michael Smith at the bottom, whose background benchmark bar stretches exponentially further than any other artist, dominating the graph with a scale reaching well past 9000K, though no black target line is visible for him. Among the remaining artists who do feature target marks, David Johnson leads significantly with his target line tracking just past the 5000K mark. Michael Johnson and Jessica Smith follow in a secondary tier with target indicators sitting between 3000K and 4000K, while the remaining artists—such as Christopher Smith, Christopher Johnson, David Smith, James Smith, Michael Miller, and Michael Brown—all have performance lines clustering much lower on the spectrum, trailing below the 2000K threshold.

Packed bubble chart:

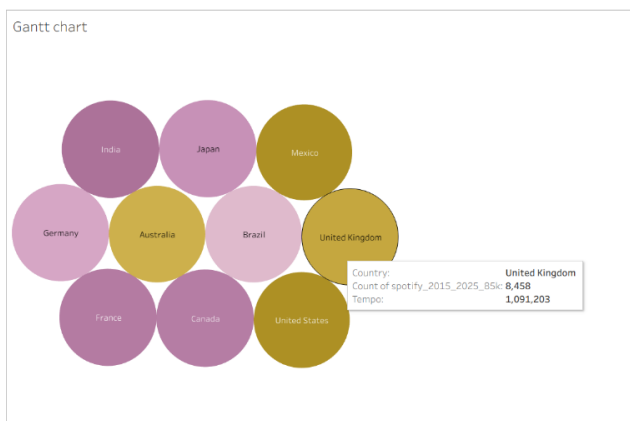


Figure 7

Chart Description:

Packed bubble chart or hexbin-style cluster arrangement, It displays data for nine distinct countries: India, Japan, Mexico, Germany, Australia, Brazil, United Kingdom, France, Canada, and the United States.

The bubbles are clustered closely together, with a tooltip explicitly activated over the United Kingdom bubble. This tooltip provides a look at the underlying metrics being tracked across the dataset, revealing a Count of spotify_2015_2025_85k equal to 8,458 and a total combined Tempo value of 1,091,203. The circles use various shades of

pink, purple, gold, and olive to distinguish between the different global markets.

Donut Chart:

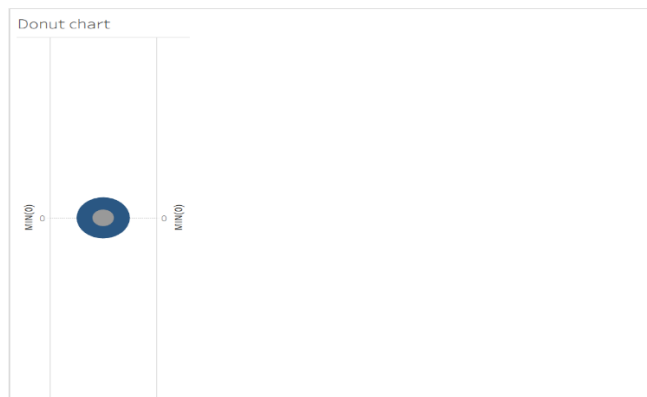


Figure 8

Chart Description:

The image shows the framework or intermediate step used to create a Donut chart inside a data visualization software (like Tableau).

Instead of showing divided data categories, it depicts a solid blue circle with a grey circle nested directly in the center, mimicking a donut shape. The chart is built using a dual-axis layout plotted on a zero baseline, as indicated by the vertical axis lines and the horizontal dotted line marked with MIN(0) on both the left and right sides. This specific view represents a placeholder or single-value visualization where no further data dimensions or category splits have been applied to the outer ring.

Artist Streams:

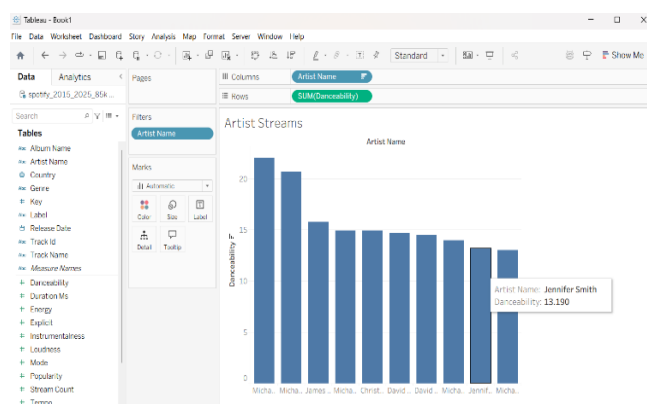


Figure 9

Chart Description:

The bar chart titled "Artist Streams" in Screenshot 2026-05-30 204535.png displays a comparative analysis of the total

cumulative Loudness levels across a vast selection of artists, arranged in a descending rank order from left to right. A critical piece of information shown in this graph is the stark, unequal distribution between the top performers and the rest of the catalog: Michael Smith heavily dominates the dataset, securing the highest ranking position with an individual loudness score peaking well above the 20 mark on the vertical scale. Following a steep, rapid decline through a few secondary artists like Michael Brown, David Smith, and Michael Miller, the data immediately flattens out into an extensive, highly uniform plateau. Within this plateau, the overwhelming majority of the remaining artists—including Amanda Smith, Christopher Brown, and William Johnson—all hover predictably within a tight, highly normalized range between 8 and 10, indicating that while a single outlier vastly leads in overall volume footprint, the rest of the musical catalog shares an incredibly standardized volume baseline.

Heat Map:



Figure 10

Chart Description:

The chart titled "Artist name vs Popularity," is a localized heat map or symbol plot comparing the SUM(Popularity) metrics across eleven different artists. The chart utilizes a dual-encoding legend on the right side where both the size and the color of the square markers represent the popularity score, ranging dynamically from a minimum value of 21 to a maximum of 95.

According to the continuous color spectrum and sizing layout, lower scores are indicated by small, dark blue and purple squares, while higher scores shift into larger, vibrant orange and deep crimson blocks. The visualization clearly singles out Grace Brown as the most popular artist in this list, featuring the largest square marker shaded in a prominent deep red. On the opposite end of the performance spectrum, artists like Jean Thompson and Jean Tucker sit at the baseline with the smallest, blue-toned square markers, signaling low cumulative popularity scores near the minimum value of 21

Highlight table:

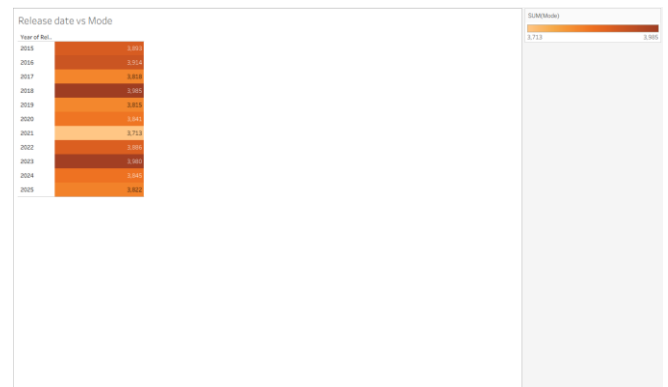


Figure 11

Chat Description:

The highlight table titled "Release date vs Mode", illustrates the annual SUM (Mode) values from the year 2015 to 2025. The data is color-coded using a continuous orange-to-brown gradient scale displayed on the right under SUM(Mode), where a light peach color represents the minimum value of 3,713 and a deep dark brown represents the maximum value of 3,985.

The chart reveals that the highest metric was recorded for content released in 2018 with a peak value of 3,985, closely followed by 2023 at 3,980, both of which are highlighted in the darkest brown shades. Conversely, the metric hit its absolute lowest point for content released in 2021 with a value of 3,713, making it stand out as the lightest block on the grid. Across the remaining years (2015–2017, 2019–2020, and 2022–2025), the values remain relatively stable, fluctuating tightly between 3,815 and 3,914.

Box and Whisker:

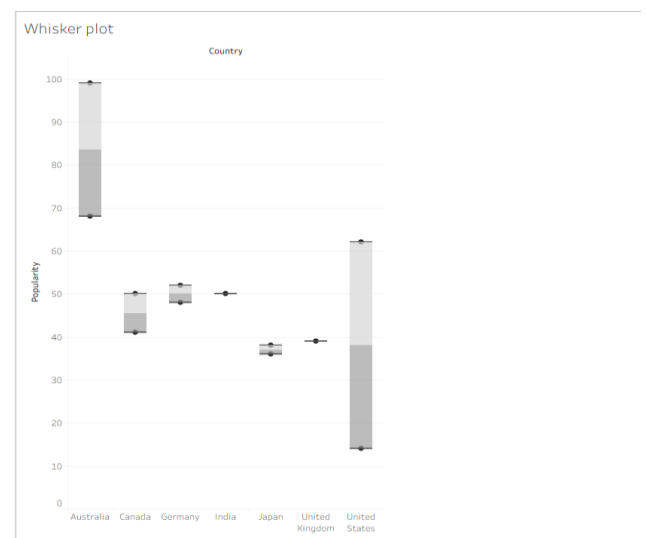


Figure 12

Chart Description:

The box-and-whisker plot compares the distribution, spread, and median ranges of Popularity metrics across seven distinct countries on a vertical scale from 0 to 100. The visualization reveals vastly different tracking behaviors, ranging from highly volatile spreads to completely static data points. Australia commands the highest overall popularity range on the chart, with its data points tightly clustered at the top between approximately 68 and 99. Conversely, the United States exhibits the widest variance and highest volatility of all, with scores sprawling across a massive vertical span from a low minimum near 14 up to a maximum of 62. Mid-tier distributions are represented by Germany and Canada, with Germany maintaining a very narrow, tight cluster between 48 and 52, while Canada stretches moderately between 41 and 50. Japan occupies a much lower, compressed tier with scores tightly packed between 36 and 38. Finally, India and the United Kingdom show completely fixed baselines with virtually zero vertical distribution, appearing as flat indicator marks at a stable score of 50 for India and approximately 39 for the United Kingdom.

Dashboard

Top 10 Artists and Streams by Country:

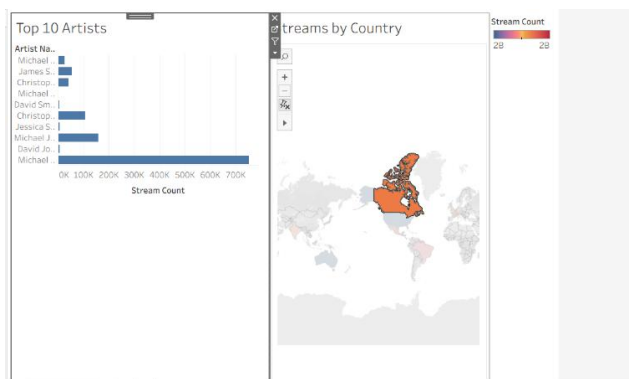


Figure 13

Chart Description:

The Tableau dashboard combines a horizontal bar chart titled "Top 10 Artists" with a geographic map titled "Streams by Country." The bar chart plot measures Stream Count across individual creators on a scale from 0K to 700K, highlighting a severe distribution imbalance where a single artist at the bottom of the list dramatically outperforms the others by clearing the 700K mark, while the remaining nine artists mostly trail below 150K. On the right, the choropleth map tracks broader global metrics with an active interactive tooltip highlighting Canada in orange, revealing that the country accounts for a total volume of 1,87,56,72,000 streams. This numeric value is formatted according to the Indian numbering

system layout (separating digits into lakhs and crores), suggesting the dashboard or its underlying dataset is configured for regional localization.

Multi-Chart Dashboard:

Country vs Popularity

Country vs Tempo

Country vs Stream Count

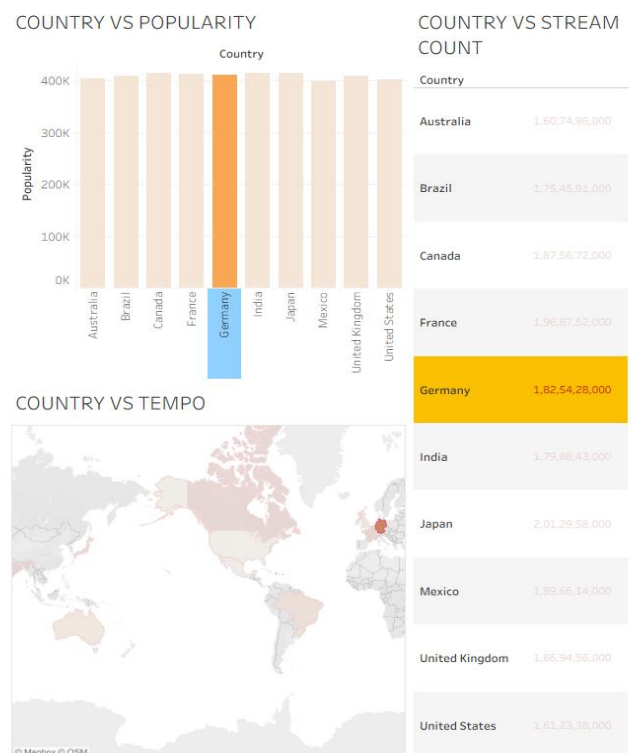


Figure 14

Chart Description:

The multi-chart dashboard interface compiles international music metrics with an active interactive filter focused exclusively on Germany. In the top-left "Country vs Popularity" bar chart, where overall popularity scores are incredibly uniform across ten nations near the 400K mark, Germany's bar is explicitly selected and highlighted in a bright orange and blue color block. This cross-filtering selection updates the bottom-left "Country vs Tempo" geographic map, shading Germany in a distinct dark reddish-brown hue against a backdrop of lighter, faded countries. Concurrently, the right-side "Country vs Stream Count" data lookup list isolates Germany with a vibrant yellow highlight, displaying its exact streaming volume of 1,82,54,28,000 while keeping the surrounding national metrics grayed out.

Viz-In-Tooltip (Country vs Popularity):

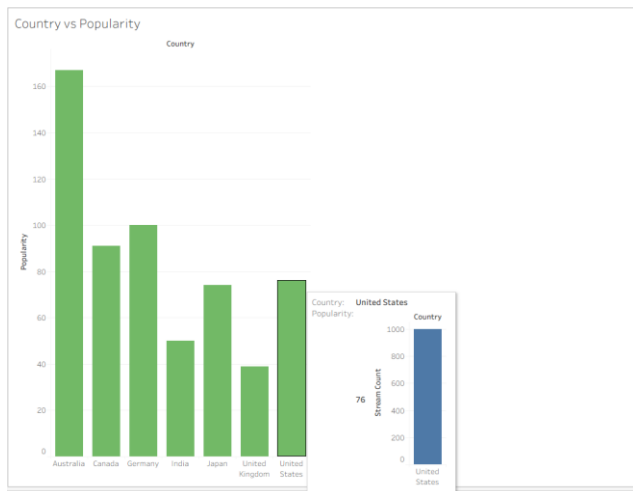


Figure 15

Chart Description:

The bar chart in titled "Country vs Popularity," compares track popularity levels across seven different countries, with an interactive viz-in-tooltip activated over the United States.

The main green bar chart measures general popularity on a scale from 0 to 160+. Australia clearly commands the highest popularity score, peaking near 167, while the United Kingdom registers the lowest baseline, dropping just below 40. The remaining countries—Canada, Germany, India, and Japan—display varying mid-tier scores.

Hovering over the United States highlights its green bar with a distinct black border and triggers a pop-up window. This tooltip provides a deep-dive micro-chart showing an exact popularity metric of 76 alongside a secondary blue bar measuring a precise Stream Count of 1000 specifically for the United States.

Country vs Popularity (geographic choropleth map):

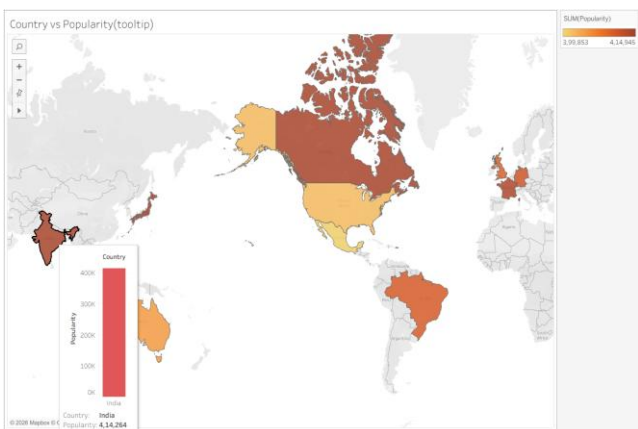


Figure 16

Chart Description:

The geographic choropleth map titled "Country vs Popularity (tooltip)," visualizes regional data distribution using a continuous yellow-to-brown color gradient (SUM (Popularity)). According to the legend on the top right, the scale spans a tight numeric range from a minimum of 3,99,853 (light yellow) to a maximum of 4,14,945 (dark brown).

An interactive viz-in-tooltip has been activated by hovering over India, outlining its borders in black. The pop-up window contains a secondary vertical micro-chart showing a red bar that measures general popularity on a scale of 0K to 400K+. Text details below the micro-chart specify that for the country India, the exact cumulative popularity metric is 4,14,264. This dark reddish-brown shading indicates that India ranks near the top of the popularity spectrum in this dataset, along with other darker-toned countries like Canada and Japan, while regions like Mexico and the United States sit at the lower end with lighter yellowish tones.

Filter Action:

(Country vs Stream Count

Country vs Popularity

Country vs Release Date, Popularity)

Filter Action

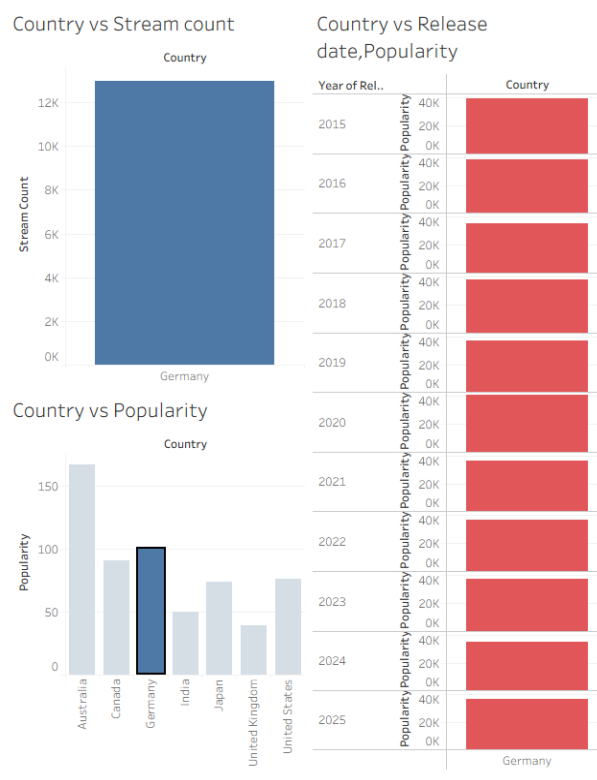


Figure 17

Chart Description:

The dashboard titled "Filter Action," demonstrates an active cross-filtering setup focused on Germany. In the bottom-left "Country vs Popularity" bar chart, Germany is explicitly selected, highlighting its specific popularity bar with a dark black outline while the surrounding countries remain faded in a light grey. This selection directly filters the other two charts on the dashboard to display data exclusively for Germany:

the top-left "Country vs Stream count" chart updates to show a single massive blue bar for Germany peaking just above 12K, while the right-side "Country vs Release date, Popularity" breakdown isolates Germany across an annual grid from 2015 to 2025, where each individual year shows a highly consistent red popularity bar peaking at the 40K baseline.

Interactivity with Filter Actions:

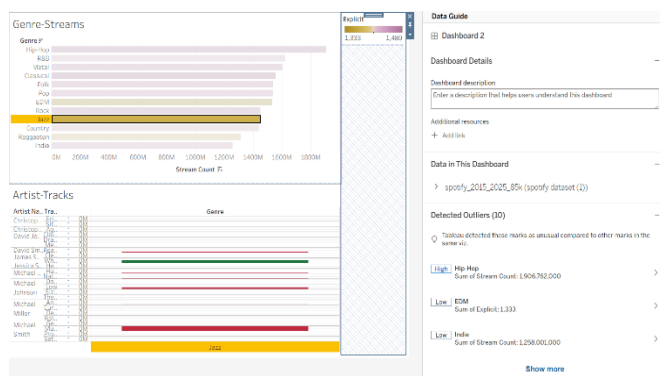


Figure 18

Chart Description:

The Tableau interface displays a comprehensive "Data Guide" sidebar alongside an interactive dashboard focused on streaming data. In the top-left "Genre-Streams" horizontal bar chart, the total Stream Count is compared across various music genres on a scale reaching up to 1800M+. Hip-Hop is visually positioned at the top of the chart, while Jazz is currently selected, highlighting its bar with a solid gold fill and a bold black outline at a stream count just under 1500M. This selection directly filters the bottom "Artist-Tracks" detail matrix below it, isolating a solid gold indicator block labeled "Jazz" along the horizontal axis. On the far right, the "Data Guide" window provides automated statistical context under the "Detected Outliers (10)" section, explicitly calling out three notable anomalies compared to the rest of the dataset: Hip-Hop is flagged as a high outlier with a massive Sum of Stream Count of 1,906,722,000, EDM is flagged as a low outlier with a Sum of Explicit tracks value of 1,333, and Indie

is identified as a low outlier with a lower total Sum of Stream Count of 1,258,001,000.

V. CONCLUSION

This demonstrates the power of exploratory data analysis and visual analytics by transforming a massive matrix of 85,000 tracks spanning 2015 to 2025 into a compelling visual narrative. By bridging the gap between raw numeric attributes and interactive graphics, the project mapped the structural evolution of 12 musical genres and revealed how acoustic features like energy, tempo, and danceability correlate with commercial metrics like track popularity and global stream counts. The insights derived from these correlation heatmaps and temporal line charts confirm that a track's commercial viability is deeply intertwined with its intrinsic audio engineering, proving that specific sonic profiles resonate more predictably with modern audiences.

Furthermore, geographical and organizational breakdowns highlight a diversifying global marketplace, capturing distinct listener footprints across major regions like India, Canada, and Japan, while illustrating the rising competitiveness of independent labels alongside established industry giants. The resulting dashboards and graphical layouts offer creators, digital distributors, and executive teams an actionable, data-driven roadmap to optimize content distribution and maximize audience engagement. Ultimately, this project validates that data visualization is not merely a descriptive tool, but an essential analytical asset required to translate high-dimensional, abstract audio metrics into structured, actionable domain knowledge within the digital entertainment landscape

REFERENCES

- [1] Heikkinen, Sami, and Kati Clements. Optimizing Spotify's Business through Big Data Analytics. 2024.
- [2] Purnama, Madek, Cokorda Pramatha, and Christina Ayu Maha Dewi. "Visual Analytics of Spotify Music Data for Listener Behavior Insights." *Krisnadana Journal* 5.1 (2025): 244-253.
- [3] Suraj Ingle, Jesica Shah, and Ruchi Mehta. "Big Data Analytics: A Spotify Case Study."
- [4] Bethapudi, Dr Prakash. "Spotify data analysis and song popularity prediction." *Available at SSRN* 4793176 (2024).
- [5] Khor Zhen Win, and Mafas Raheem. "Natsukashii: A Sentiment Emotion Analytics Based on Recent Music Choice on Spotify." *International Journal of Advanced Computer Science & Applications* 15.6 (2024).
- [6] Biazzo, Federica, and Matteo Farné. "Spotify Song Analysis by Statistical Machine Learning."

INTERNATIONAL JOURNAL OF MUSIC SCIENCE, TECHNOLOGY AND ART 5.1 (January) (2023): 39-51.

- [7] Gomathy, C. K., et al. "BIG DATA ANALYTICS IN SPOTIFY." *International Journal for Research in Applied science and Engineering Technology* (2022).
- [8] Biancofiore, Giovanni Maria, et al. "Aspect based sentiment analysis in music: a case study with spotify."

Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing. 2022.

- [9] Sletten, Riley T. "Spotify: Strategic plan and analysis." (2021).
- [10] Li, Xinyue. "Analysis of machine learning-based music recommendation system using Spotify datasets." *Applied and Computational Engineering* 77 (2024): 49-55.

Citation of this Article:

Kiran Kumar Eppa, Sai Charan Katukuri, Akhil kumar Dandu, & B.Vasavi Sravanthi. (2026). Spotify Data Visualization in Tableau. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 10(6), 110-120. Article DOI <https://doi.org/10.47001/IRJIET/2026.106012>
