**2024 Music Trend Analytics Report**

Data 1101, Intro to Business Analytics

Dr. Kanlun (Sampson) Wang

Due by Friday., Dec. 6, 11:59 PM

**Group Leader:** Kevin Hanson

**Group Members:** Andrew Loughlin, Katherine Jakubowsky, Michael Lye

**Table of Contents**

Introduction (Micheal Lye)

* A summary of the project background
* What specific business/research questions does your group aim to address?

Methodology (Kevin Hanson)

* **Data collection:** please ensure to appropriately cite the data source and discuss some important variables/features if not all.
* **Data pre-processing:** please thoroughly discuss how you pre-process the collected data, including but not limited to missing value treatment, data manipulation, and data transformation.
* **Data analysis:**
  + Please clarify the input variables and/or target variables within the context of this project. If applicable, multiple input variables and/or target variables may be involved.
  + You can utilize any techniques you have acquired during this course, including pivot tables, descriptive statistics, normality testing, regression, and their combinations.
  + In addition, you have to provide strong arguments for your method selection(s).
* **Evaluation:** If applicable, please describe your plan for evaluating the predicted outcomes, including the utilization of some metrics, if not all.

Results (Katie Jakubowky)

* Key findings
* Please present some visualizations of your outcomes, such as using tables, figures, and charts, and provide a comprehensive discussion of the significant results for each data analysis output.

Conclusion (Andrew Loughlin)

* Discuss whether the research/business questions have been successfully addressed by summarizing the major findings.
* Contributions or implications to real-world business scenarios or others.

References

**Introduction**

From decade to decade the music industry is constantly changing in response to new sounds, trends, and platform popularity. As these shifts occur new artists, labels, and opportunities open as the industry is disrupted by constantly shifting preferences. From disco in the 1980s to the country’s rising popularity in recent years, music trends dictate what dominates the charts and influences consumer behavior. The purpose of this project was to explore key factors such as artist dominance, trends in release dates, and the role of streaming platforms in driving engagement and trends. By leveraging the power of data analytics, this study aimed to answer essential research questions about the influence of trends on the music industry and provide actionable insights that could benefit artists, record labels, and streaming platforms. The analysis connects to the course by utilizing skills learned in Python and Excel to normalize, analyze, and compartmentalize data into meaningful insights.

Our research questions, aimed to pinpoint the driving factors pushing the trends to clearly define where the future of the decade is going. Our first question dove into the top tracks using a combination of factors including track rank, total streams and views, and playlist counts. This would give us the top-performing class across all platforms and the nature of these songs illustrates trends. Our second question dove directly into the center of our research outlining the top platforms for the entire year revealing “where” the top tracks were being played and the trends the trends were formed. Finally, we completed our research questions defining the top performing artists of the year as well as the popularity of songs released in recent years compared to the past. These research questions outlined out overall goal to pinpoint the driving factors and clues to the musical iindustry'strends for the rest of the decade.

**Methodology**

After defining our research questions and the objectives of our research we dove into Kaggle to find a dataset that would provide us the information we would need to answer each of our research questions. After exploring the platform we eventually came across a dataset titled, “Most Streamed Spotify Songs 2024” by Nidula Elgiriyewithana. We were originally drawn to this dataset because of its popularity on Kaggle as well as a diverse range of data to answer our research questions. Before exploring the data within Python we evaluated the description given on Kaggle giving insight into what the dataset included, potential drawbacks, and analysis by previous community members. During our evaluation, we discovered that the dataset had a few flaws including null values as well as a lack of updates. With this in mind, we deduced that we would analyze the data that was complete as well as date our analysis back to the time of the last update.

After evaluating and understanding and dataset we began to analyze the data in Python using numerous libraries including, pandas, numpy, and matplotlib.pyplot to manipulate, process, and analyze the data. After exploring the datasets column, rows, and data types we found that some of the null values would need to be removed as well as various data types normalize to complete our analysis. To normalize the data we changed object data types to int as well as replaced commas with spaces for all of the numerical values to be accurate for data analysis. Additionally, some columns including “TIDAL Popularity” and “Explicit Track” were dropped because of the lack of relevance to our research questions.

Once the data was normalized we began to answer our research questions using pandas to combine and divide columns within the data frame. The first column we added named “Total Streams & Views” included all of the columns that related to either streams or views across all platforms. The next column followed the same process but added all of the columns that were related to playlists across all platforms. After adding these two columns we were able to answer our first research question which illustrated the top-performing tracks by rank, streams and views, and playlists across all platforms.

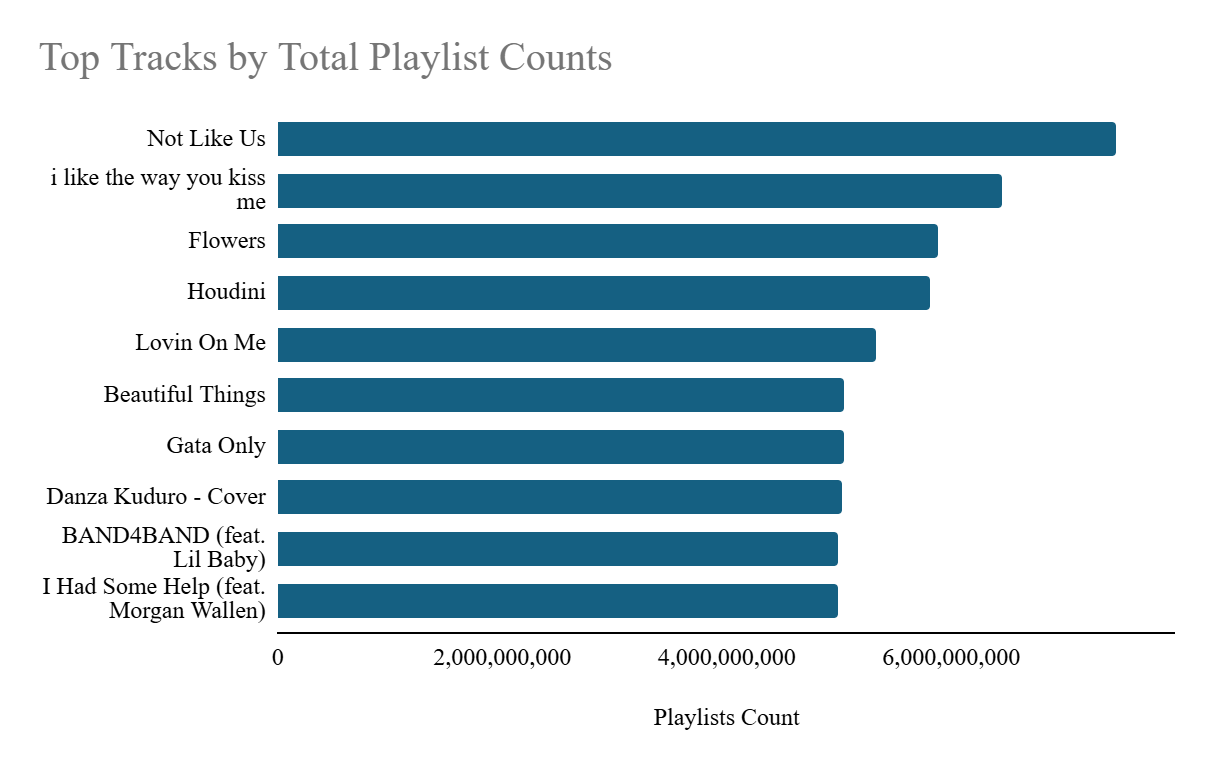
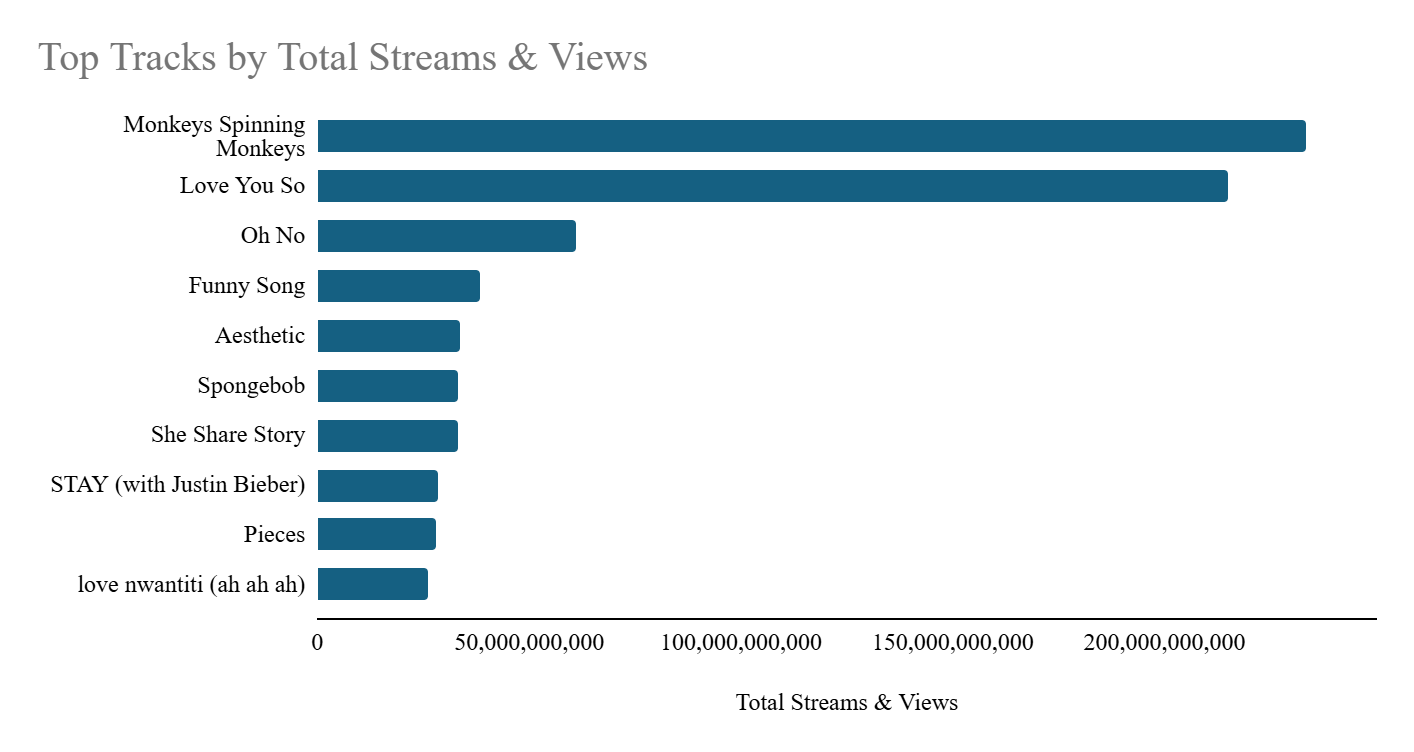
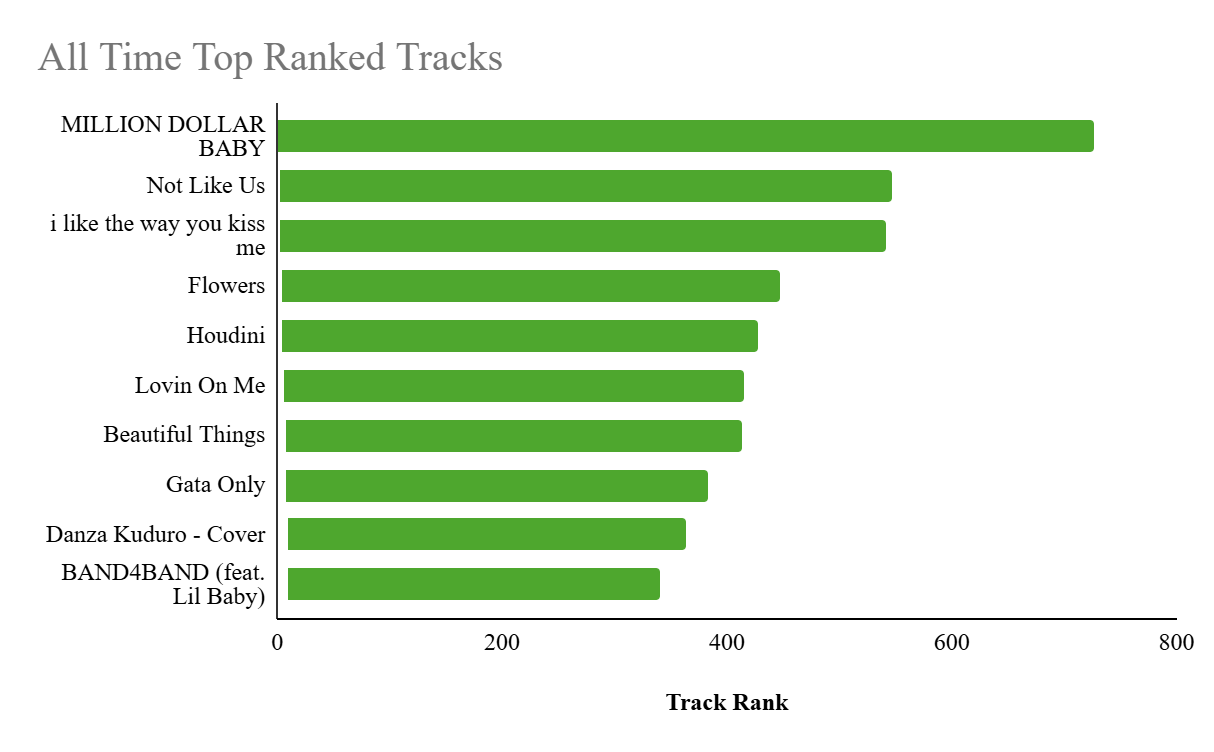
For the second question, we followed a similar process grouping all of the platforms and adding a new column for each. This allowed us to find the most popular platforms for the year giving a deeper insight into forming trends. After completing the first two questions using pandas we moved into a statistical analysis of the data to answer our final research question.

Finally, for the last research question we decided to use numpy and the statistics libraries and analyze the top artists and songs by release date for our last research questions. To answer our first question we found the sum of tracks for each artist and then created a data frame that described the amount of tracks each artist had with the dataset. Similarly, for a release date for each track, we created a final data frame that counted the number of songs released during different years.

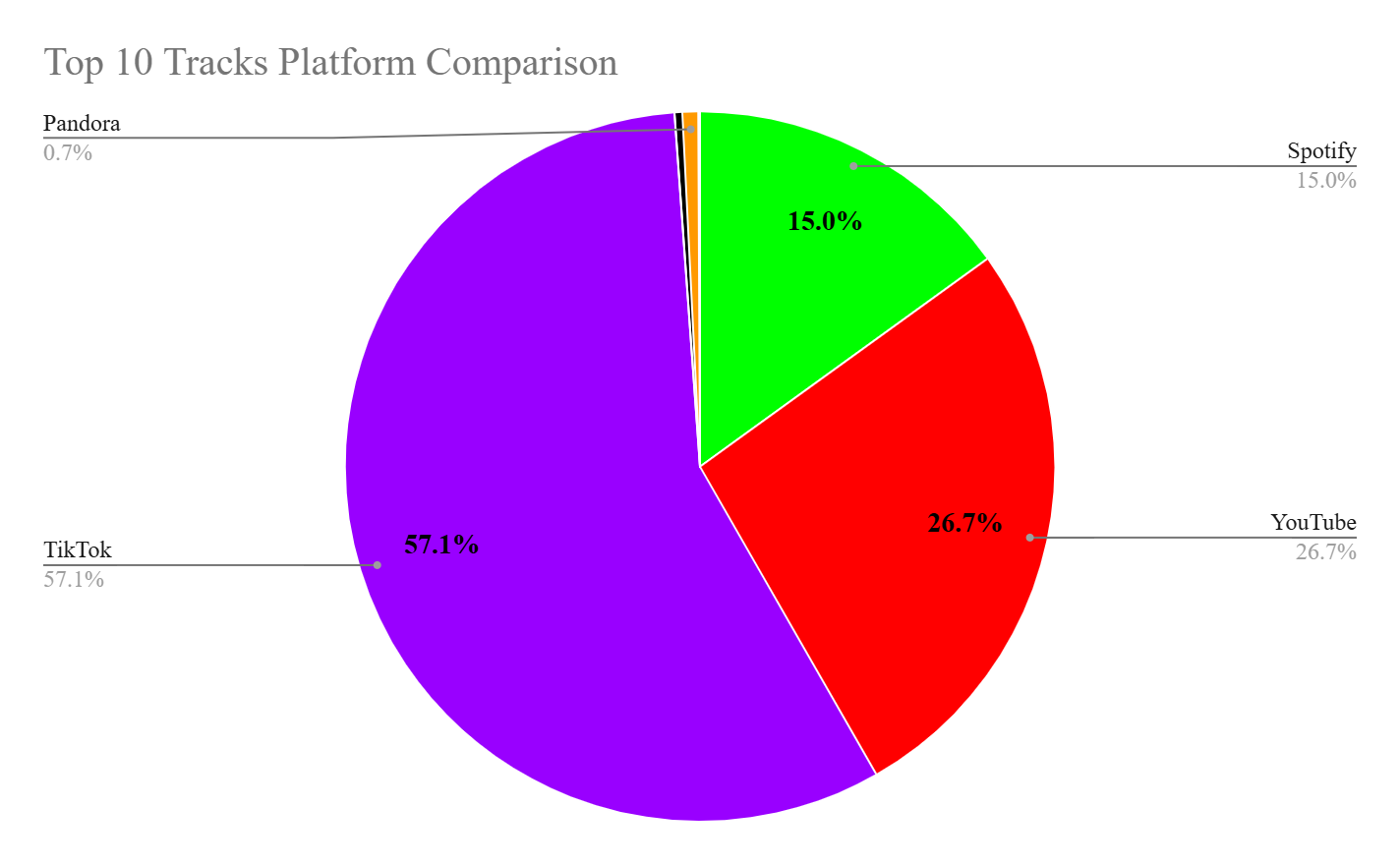
After processing, normalizing, and analyzing our data we created three separate data frames related to each of our research questions and exported them to Excel to interpret and visualize our analysis of the dataset.

**Results**

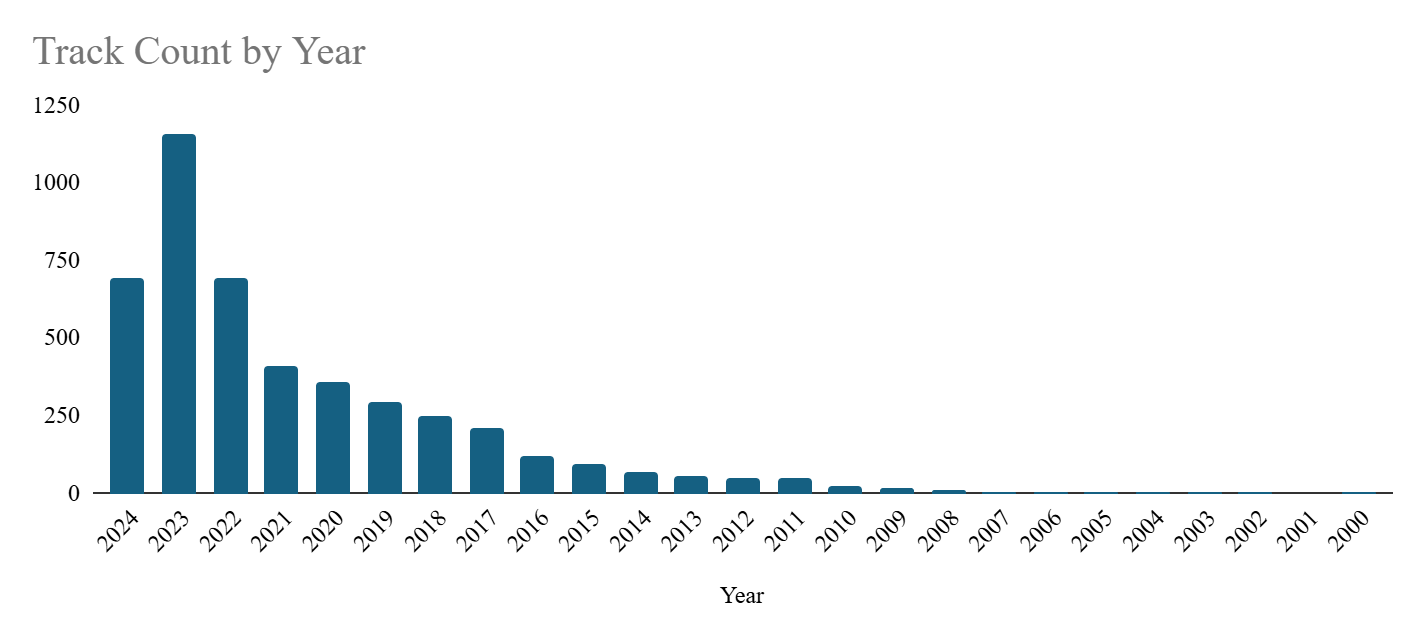
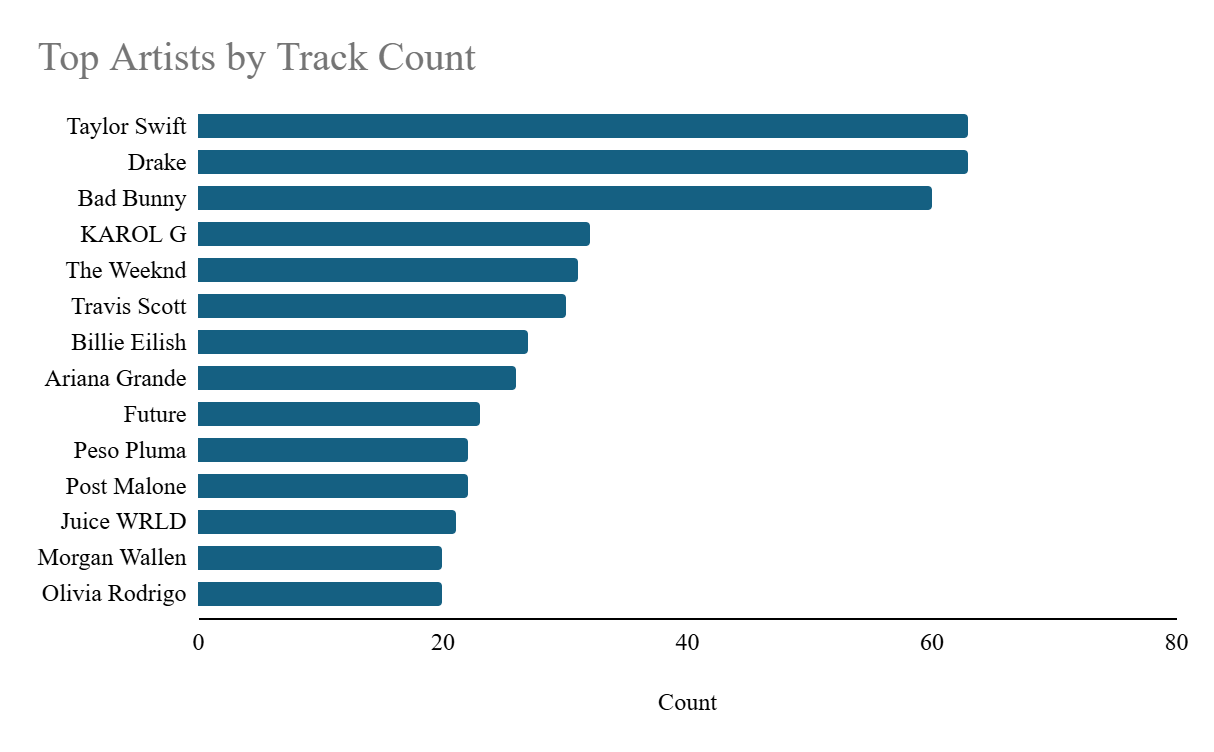
Moving out of Python and into Excel we easily were able to format and visualize each data frame to answer each of our research questions to pinpoint the key drivers of emerging trends for the rest of the decade. For the first research question, we decided to sort the data to showcase the top 10 songs for track rank, stream and views, and playlist counts. After sorting our data we created three different horizontal bar charts that showcased the top-performing tracks that dominated the first half of the year. After creating this visualization we revealed a strong correlation between the tracks that appeared at the top for track rank, total streams and views, and playlists. With this in mind, we also didn’t recognize a majority of the tracks at first but after listening to each of them we found that a majority of the top-performing songs were from major TikTok trends suggesting a direct correlation between TikTok and top-performing tracks for the year. **(Visualization shown below)**



With these insights in mind, we completed a bar chart visualization to answer our second questions revealing once again that the top platform for the year was TikTok. This was showcased by TikTok being the most popular platform at 57.1% followed by YouTube at 26.7% and Spotify at 15%. After completing this visualization we reinforced our original hypothesis that the data would be skewed by the recent popularity of the TikTok platform. **(Visualization shown below)**



Finally, our last research question aimed to validate our hypothesis that the nature of top-performing tracks reflects emerging trends in the music industry. To find these insights we returned to our original visualizations with a horizontal bar chart to display top-performing artists and a vertical bar chart to showcase the count of tracks by release date. After completing this visualization our hypothesis stood strong with more songs released in recent years showing up in the dataset. Additionally, the top artists fell in line with our expectations illustrating incredibly popular well-known artists such as Taylor Swift and Drake appearing at the top of the list. **(Visualization shown below)**



**Conclusion**

This research delved into the dynamic landscape of the music industry, aiming to uncover key factors driving trends and influencing consumer behavior. By analyzing a comprehensive dataset, we sought to address specific research questions related to top-performing tracks, popular platforms, and the impact of release dates on song popularity.

**Major Findings**

* **TikTok's Dominance:** The study underscored the immense influence of TikTok on the music industry. The platform emerged as the dominant force, contributing significantly to the popularity of tracks and shaping emerging trends.
* **Correlation Between Platforms and Track Performance:** A strong correlation was observed between the performance of tracks across various platforms. Top-performing tracks consistently excelled on multiple platforms, indicating a synergistic relationship between streaming services and social media.
* **Recent Releases and Trendsetting:** The analysis highlighted the impact of recent releases on the music industry. Newer songs were more likely to appear in the dataset, suggesting a preference for contemporary music and the rapid evolution of trends.

**Implications for the Music Industry**

* **Artists and Record Labels:** Understanding the role of platforms like TikTok can help artists and labels strategize their marketing and promotional efforts to maximize exposure and fan engagement.
* **Streaming Platforms:** The study emphasizes the importance of fostering a diverse ecosystem of platforms to cater to different listener preferences and emerging trends.
* **Music Industry Analysts:** The findings provide a data-driven framework for analyzing music trends and making informed predictions about future industry developments.

By leveraging data analytics and insights from this research, industry professionals can make informed decisions to navigate the ever-evolving music landscape and capitalize on emerging opportunities.

**References**

*Most Streamed Spotify Songs 2024*. (2024, June 15). Kaggle. https://www.kaggle.com/datasets/nelgiriyewithana/most-streamed-spotify-songs-2024

Lance, M. (2024, December 2). *Minimal Pop Music Festival. PPT & Google Slides Template — SlidesCarnival*. SlidesCarnival. https://www.slidescarnival.com/template/pop-music-festival/56434

U, V. (2022b, January 4). My Experience from 3 years of Kaggle - Vishnu U - Medium. *Medium*. https://vish0399.medium.com/my-experience-from-3-years-of-kaggle-a4c4d8d68974