The Social Soundscape: How Social Media Impacts Music Popularity

Group 20: Derek Wang, Henna Abassi, Sarina Doss, Maeve Horan-Portelance, Aidan Perez, Naomi Suzuki, Sydney Tomsick

Abstract

As social media platforms and algorithmic content feeds grow in popularity, we are curious as to how they may influence our music listening habits. While previous research on musicality and trendiness exists, we specifically aim to analyze how different social media platforms and trends may alter the listening habits of the public, whether through recommendation algorithms prioritizing certain release schedules, selective promotion of certain genres on certain platforms, or even cross-platform popularity trends. We utilize various statistical methods to achieve conclusions based on our dataset of tracks and their social media performance. We find that although there are clear links between social media performance and song popularity, there is still much more research to be conducted on the quantification of these effects. Our conclusions seek to both inform the public on the influence of social media platforms and algorithms on our music listening and also to possibly improve and/or ameliorate social media platforms' decisions in advertising and recommendations.

Introduction

The growth of social media has invaded fundamentally every facet of our lives and music is no outlier. With algorithmic feeds and sharing from our mobile devices, the internet and social platforms greatly impact what songs are in the current zeitgeist. In this project, we aim to analyze how social media impacts music popularity. We additionally aim to quantify how much social media affects track popularity, and how different platforms may vary in their influence. Our goal is to discover and inform our peers of how much social media may be affecting our individual lives and choices, and how people could be more mindful of music consumption.

Literature Review

There has been a plethora of research on "hit song science": using data science to predict which songs will be most popular. Much of this research focuses on how the sound of the song (acousticness, danceability, energy, speechiness, tempo, instrumentalness, etc.) affects its popularity. For example, a study (Middlebrook & Sheik, 2019) used Spotify data with many of these factors (as well as a few others such as artist name, whether the song was explicit or not, duration, and album release date) to predict the Billboard top 100 songs between 1985 and 2018. The authors used logistic regression, neural networks, random forest, and support vector machines to find the best-performing model that accurately predicted which songs were the most popular based on the given factors. Other articles such as “Spotify: You Have a Hit!” (Dawson et al., 2021) and “Predicting song popularity in the digital age through Spotify’s data” (Li, 2024) also use Spotify data to identify which sound-based factors to identify have the greatest association with song popularity.

With the rapid growth of social media over the past decade, particularly the rise of TikTok, social media has become an “essential promotional tool” (MediaTech, 2023) in the music industry. A recent article by MediaTech highlights TikTok's significant role in modern song promotion. The article explains how marketing teams strategically collaborate with influencers to feature songs, fostering organic engagement. This subtle form of advertisement makes it less apparent to fans that a song is being promoted, enhancing its authenticity and reach. By leveraging viral sounds, this promotional strategy creates a mutually beneficial relationship between artists and social media platforms. Artists gain increased visibility, engagement, and streaming numbers, while platforms like TikTok benefit from higher user interaction and prolonged app engagement.

Although there is a lot of research on whether sound-based factors influence song popularity, there is much less statistical research on the connection between Spotify song popularity and other social media and music streaming platforms. Our team wants to discover how TikTok and YouTube engagement affects song popularity and how Spotify song rankings/streams differ from Apple Music song rankings/streams. There is some research on TikTok’s impact on the re-popularization of older songs (Matos et al., 2024). There are also articles written about TikTok’s effect on music popularity (Jorgenson, 2022) that approach the topic from a humanities perspective. However, there are very few data-based studies on this topic. We hope to fill this gap in research through our project.

Data Collection and Cleaning

We used a dataset from Kaggle entitled “Most Streamed Spotify Songs 2024”. It contains the title, album name, artist name, and release date of the most streamed songs on Spotify in 2024. It also contains the number of streams, likes, views, and/or posts on various music streaming and social media platforms such as Spotify, Youtube, TikTok, Apple Music, Pandora, Soundcloud, Shazam, and radio stations.

We cleaned the data using R to remove duplicates, remove unnecessary columns, standardize the release date column, and create a column called “Release.Month” to answer our first research question. Further, the columns that contained numbers of streams/likes/posts/views were originally stored as categorical variables and contained commas within the numbers. We eliminated the commas and made these variables numeric.

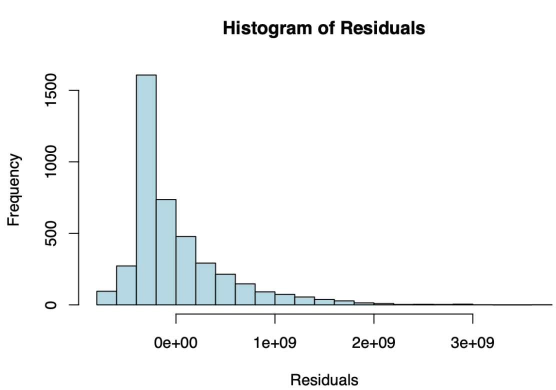
Lastly, there were many NA values in our dataset. First, we removed the rows that contained NA values in the columns “Spotify.Streams”, “YouTube.Views”, or “YouTube.Likes”. We did this because these are vital variables for our research questions, and it only removed a small proportion of observations (we lost 317 rows, but there are still over 4000). Next, we filled the rest of the NA values with 0. We did this because, when going through the data and cross-checking with streaming/social media platforms, most of the NA values represented a very low number of streams/likes/posts/views, if any at all. For example, if a song had a missing value for TikTok posts, that song usually had 0 TikTok posts.

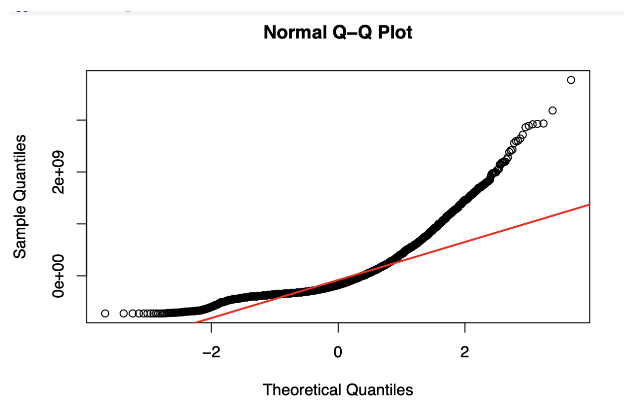
Research Question 1

Does the release month of a song impact its popularity?

Methods

Initially, we planned to use a one-way ANOVA to compare the mean of Spotify streams across different months. However, after checking our assumptions for ANOVA, we realized that the normality assumption was violated and we needed to find a different method. First, we checked the normality assumption using the Shapiro-Wilk test. This test returned a very small p-value ($p < 2.2e-16$), indicating that the residuals are not normally distributed. We also checked the assumptions through residual analysis:





Our histogram of residuals and a Q-Q plot visually confirmed that the residuals were heavily skewed and did not follow a normal distribution. The histogram was not bell-shaped and dramatically skewed to the right, and the Q-Q plot showed deviations from a straight line.

Since our normality assumption is violated, instead of ANOVA, we used the nonparametric Kruskal Wallis test which evaluates whether at least one month has a statistically different distribution of Spotify streams compared to others. For the post-hoc analysis, instead of using Tukey's HSD test, we used the Wilcoxon Rank-Sum test with Bonferroni correction. Since the Kruskal-Wallis test only indicates that at least one month differs significantly in Spotify streams, the Wilcoxon Rank-Sum test evaluates which specific months are different from each other. It does this by performing pairwise comparisons: comparing the distribution of Spotify streams between each possible pair of months to determine if there is a significant difference.

Results

The Kruskal-Wallis test yielded a highly significant result (p = 2.2e-16), confirming that at least one month has a statistically different distribution of Spotify streams compared to others. This tells us that it is worth looking into further and that finding out which months are significantly different, and if they see higher or lower streams on average, could provide valuable insights.

```
##  Pairwise comparisons using Wilcoxon rank sum test with continuity correction
##
## data:  data$Spotify.Streams and data$Release.Month
##
##           January February March   April   May     June    July    August
## February  < 2e-16 -        -       -       -       -       -       -
## March     < 2e-16 1.00000  -       -       -       -       -       -
## April     < 2e-16 1.00000  1.00000 -       -       -       -       -
## May       < 2e-16 1.00000  1.00000 1.00000 -       -       -       -
## June      2.4e-12 1.00000  0.27028 0.01772 0.04979 -       -       -
## July      2.1e-06 0.00928  0.00014 1.9e-06 6.6e-06 1.00000 -       -
## August    3.5e-08 0.44925  0.01582 0.00076 0.00128 1.00000 1.00000 -
## September 5.7e-13 1.00000  1.00000 0.31343 0.43197 1.00000 0.30128 1.00000
## October   9.3e-11 1.00000  0.04943 0.00235 0.00337 1.00000 1.00000 1.00000
## November  1.4e-05 0.00686  0.00010 1.3e-06 3.8e-06 1.00000 1.00000 1.00000
## December  2.4e-10 1.00000  1.00000 0.38800 0.51301 1.00000 1.00000 1.00000
##           September October November
## February  -         -       -
## March     -         -       -
## April     -         -       -
## May       -         -       -
## June      -         -       -
## July      -         -       -
## August    -         -       -
```

Figure 3: Pairwise Wilcoxon Test P-Value Matrix

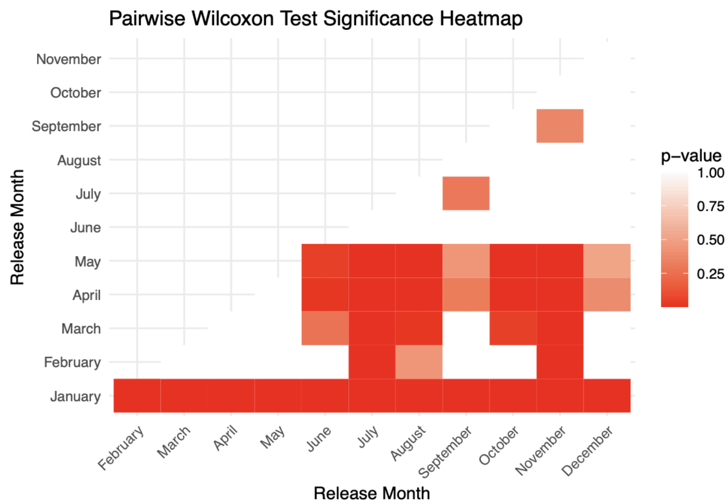


Figure 4: Heatmap of Pairwise Wilcoxon Test Significance

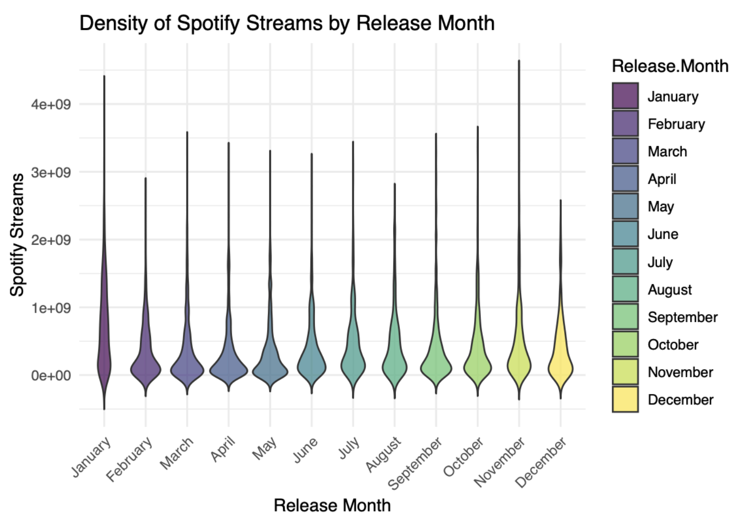


Figure 5: Violin Plot Showing Density of Spotify Streams per Release Month

Further pairwise comparisons using the Wilcoxon Rank-Sum test with Bonferroni correction provided more granular insights into which months differed significantly. This test compared each pair of months, and if $p < 0.05$, we considered the months to exhibit significantly different average number of streams. January emerged as the most distinct month, showing significantly higher streaming numbers compared to all other months, as seen by the deep red line for January in Figure 4 (significant p-value for Wilcoxon Rank-Sum Test), as well as the distribution in Figure 5, which shows the higher average number of streams. This trend could be attributed to post-holiday streaming boosts, New Year playlist curations, or reduced competition from major end-of-year releases.

February through May exhibited similar streaming patterns (no significant p-values indicated by Figure 3), suggesting a stable and consistent trend in song performance during these months. This may indicate that listener behavior during this period is not strongly influenced by seasonal shifts or external factors. July and November, however, deviated from this pattern, showing significant differences from months like March, April, and May. These variations could be explained by seasonal listening habits, such as summer vacations influencing July streams and pre-holiday anticipation driving increased engagement in November.
Interestingly, August, September, October, and December did not show significant differences from one another, implying a relatively stable streaming pattern in these months. This suggests that songs released in these months do not experience the same dramatic shifts in popularity observed in January, July, and November.

Overall, these results highlight that release month plays a significant role in determining a song's streaming performance. The p-value matrix, heatmap visualization, and violin plot confirm these findings, particularly highlighting January's distinctiveness and the seasonal variations observed in July and November. Artists and music industry professionals could leverage these insights to strategically plan release dates, maximizing exposure and engagement during months with historically higher streaming activity. Artists and music industry professionals could leverage these insights to strategically plan release dates, maximizing exposure and engagement during months with historically higher streaming activity.

Research Question 2

Are certain genres of songs more popular on certain platforms?

Data Collection and Processing

Our main dataset did not have any information on genres. As a result, we used Chordic's Spotify Playlist Analyzer to extract the parent genre of songs, producing a systematic random sample of 200 songs, 50 for each of the following genres: Pop, Hip Hop, Country, and Latino. Given that this tool only works on playlists, we selected two Spotify playlists for each genre: the top search result[1] for "Top *Genre Name* Songs of 2024" and "Top *Genre Name* Songs of 2025."[2] The Chordic tool outputted a CSV file with a number of different columns (the only one we were interested in was "Parent Genre").

In order to process the new data, we combined the playlist CSV's by parent genre (for a total of four) and then separately joined the resulting tables with our main dataset. This

[1] The exact names of the playlists used are on the .Rmd file.
[2] We combined two playlists per genre in order to get at least 50 songs. However, the top search result for Pop had over 300 songs, so we only used one playlist for Pop.

automatically removed songs from the playlists that were not in our main dataset. From there, a number of cleaning steps took place: removal of songs whose genre was not correspondent to the playlist genre, removal of repeats, removal of any songs that did not have data on TikTok, and manual removal of songs that were joined with our dataset incorrectly due to the song having a common name. We then took the top 50 songs of each based on their total YouTube, Pandora, and Spotify streams (TikTok was excluded because it does not have a music streaming sector of its platform). This resulted in 4 tables with a systematic sample of 50 of the top Pop, Hip Hop, Country, and Latino songs.

Methods

We conducted a repeated measures ANOVA to examine the effect of music genre and streaming platform on total streams of popular songs. The design included four genres (Hip-Hop, Country, Latino, and Pop) as a between-subjects factor, with 50 songs randomly sampled from each genre. Streaming platform (Spotify, YouTube, TikTok, and Pandora) served as the within-subjects factor, as each song had stream counts for all four platforms. This balanced design allowed us to compare how platform performance varied across genres and to test for possible interaction effects.

Initially, we planned to use repeated measures ANOVA to analyze these effects. However, due to significant violations of normality and variance assumptions, we opted for a non-parametric ART ANOVA instead. This method allowed us to properly assess the impact of both genre and platform, as well as their interaction, without the biases introduced by the dataset's skewed distribution.

To further explore the results, we conducted a post hoc analysis test with the Bonferroni correction method to account for multiple comparisons. These results can be split into two main categories. Posthoc_same_platform compared different genres within the same platform. This is particularly valuable for advertisers, as it highlights which genres are most successful on each platform. For example, Country music performed significantly better than Hip-Hop and Latino on Pandora, indicating that advertisers focusing on Pandora would likely achieve the best results by promoting Country music. Since platform usage varies across genres, this comparison helps advertisers make more strategic decisions about where to allocate their budgets for maximum effectiveness.

The second category is highlighted by posthoc_same_genre. This compared the same genre across different platforms, providing insights for streaming services. If a genre performed better on one platform than another, such as Country music being more popular on Pandora than Spotify, platforms could use this information to refine their strategies. For instance, a platform where a genre underperforms might work to attract more listeners from that demographic, while one where a genre thrives could double down on content and partnerships that cater to those listeners. Understanding these platform and genre relationships allows streaming services to optimize their playlist curation, and promotional strategies to better serve their audience.

Results

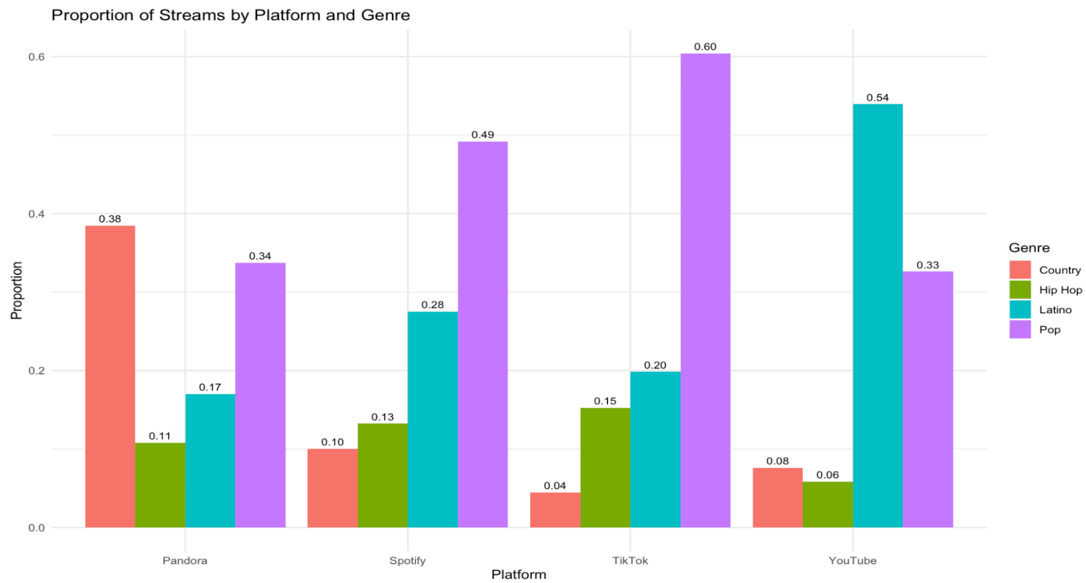


Figure 6: Proportion of Streams by Platform and Genre

This bar chart illustrates the proportional distribution of streams across platforms for each genre. The key takeaway is that certain genres dominate specific platforms. For example, Country music has the highest proportion of streams on Pandora (0.38), while Pop dominates Spotify (0.49) and TikTok (0.60). Latino music performs exceptionally well on YouTube (0.54), reinforcing the idea that different platforms cater to different audience preferences. Since the sum of each platform's bars equals 1, these values reflect relative popularity rather than absolute streaming numbers.

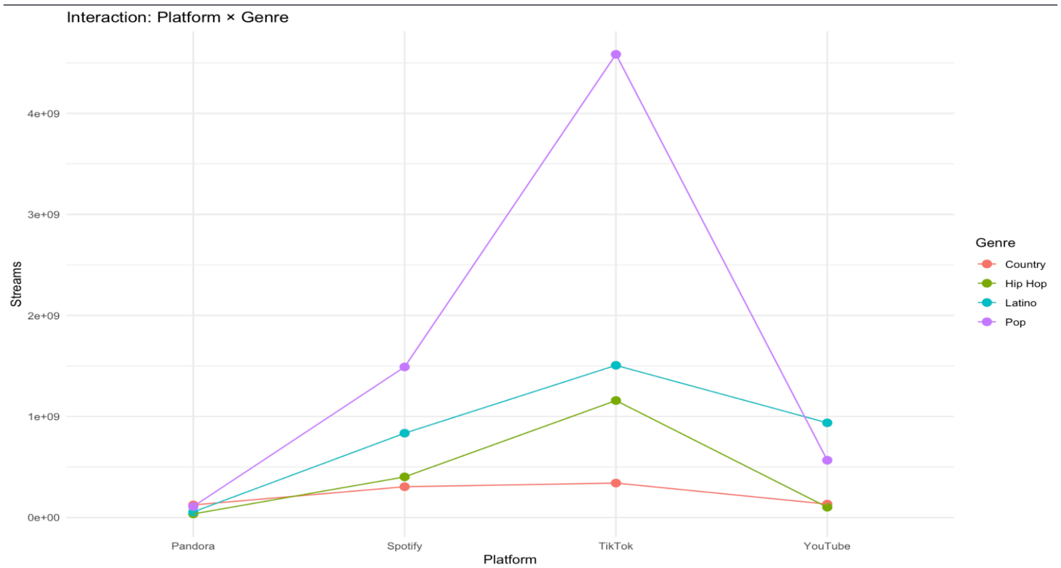


Figure 7: Interaction Between Platform and Genre

This line plot visualizes the interaction effect between platform and genre, showing how each genre's total streams vary across platforms. The increase in Pop streams on TikTok highlights its high presence on that platform, whereas Country remains relatively low across all platforms except Pandora. The intersection and divergence of genre lines suggest that the relationship between platform and genre is non-uniform, which aligns with the interaction effect found in the ART ANOVA.

```
Analysis of Variance of Aligned Rank Transformed Data

Table Type: Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger df)
Model: Mixed Effects (lmer)
Response: art(Streams)

                       F Df Df.res     Pr(>F)
1 Platform       407.242  3    588 < 2.22e-16 ***
2 Genre          134.524  3    196 < 2.22e-16 ***
3 Platform:Genre  54.614  9    588 < 2.22e-16 ***
---
Signif. codes:   0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 8: ANOVA Table

The ANOVA table output confirms highly significant effects for platform, genre (which are not important in answering the research question) and their interaction ($p < 0.0001$ in all cases). This statistical confirmation validates the observed trends in Figures 1 and 2, reinforcing the conclusion that music consumption patterns vary significantly based on the interaction between platform and genre, and strategies for marketing and content promotion should be adjusted accordingly.

Research Question 3

Is there a correlation between a song's popularity on TikTok vs. Spotify?

Methods

We chose to use a linear regression model to observe the relationship between TikTok Views, Posts, and Likes on a song's Spotify streams. After running multiple linear regression on these factors, we found that the R^2 was extremely low (0.03) and the residual error was extremely high (>500 million). In addition, the results showed that TikTok views had a negative impact on Spotify streams which demonstrated that the model was not a good fit and suggested there was a missing variable or multicollinearity.

Through checking the correlation and multicollinearity, we found that both TikTok Likes and Views have high degrees of multicollinearity and are highly correlated. As a result, we chose to remove the likes variable before running the model again.

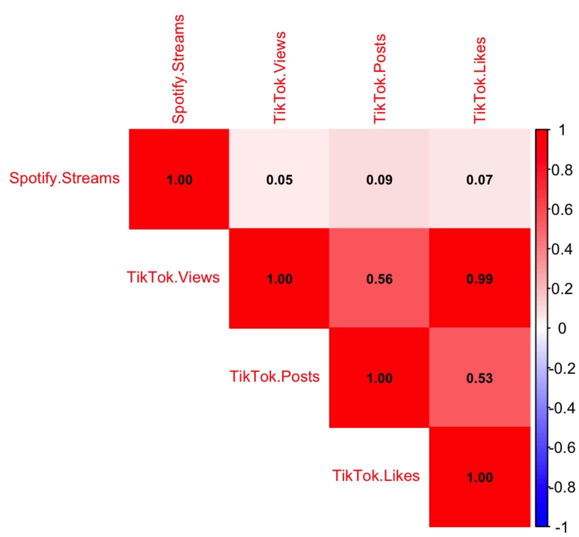


Figure 9: Correlation Matrix

Views	Likes	Posts
79.216360	1.531105	76.058724

Figure 10: Degrees of Multicollinearity

Despite this, the second model had an even lower R^2 of 0.009 and a high residual error suggesting that it was still not a good fit. Thus, we transformed the data and used a logistic regression model to evaluate the relationship between Spotify Streams and TikTok engagement.

Results

Our final model was a logistic regression with an adjusted R^2 of 0.154. This means that only about 15.4% of the variation in Spotify streams could be explained by TikTok posts and views. While this isn't particularly strong, it still gave us some interesting insights.

One of our most significant findings was that log_TikTok.Posts emerged as a highly significant predictor. According to our model, an increase in TikTok posts corresponds to an average increase of roughly 32.62 million Spotify streams, all else held constant. This suggests that songs that are posted more frequently on TikTok do tend to gain more traction on Spotify to some degree.

We also found a low overall p-value for the regression, which indicates that at least one of our predictors significantly explains some of the variability in Spotify streams. When visualizing the relationship between TikTok posts and Spotify streams, we observed a subtle upward trend in the red regression line which shows a sign of a weak positive correlation between the two.

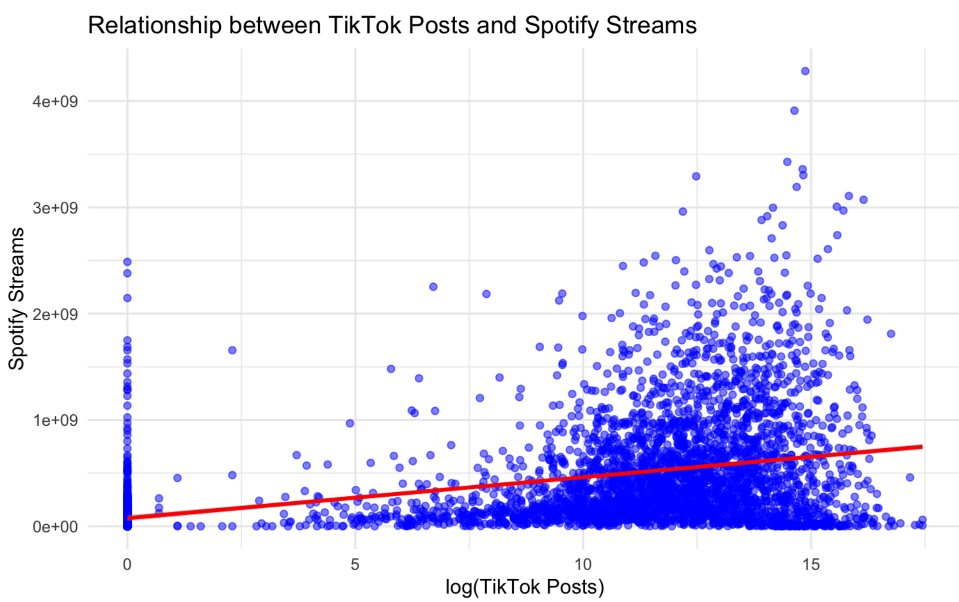


Figure 11: Relationship between TikTok Posts and Spotify Streams

However, even with transformations and removing highly collinear variables, our model was still limited. The relatively low adjusted R^2 and high residual error point to the fact that other important variables likely play a role in determining Spotify stream counts that were not captured in our data set.

Overall, despite finding some statistical significance, we ultimately concluded that there's little evidence for a strong correlation between a song's popularity on TikTok and its popularity on Spotify. While virality on one platform might boost performance on the other, it's clear that TikTok engagement alone doesn't explain most of what makes a song blow up on Spotify.

Limitations

While this study provides meaningful insights, there are several limitations to consider.

First, TikTok is not a traditional music streaming platform, so its data behaves differently from platforms like Spotify, YouTube, and Pandora. Some songs that are highly popular on streaming services may have little to no presence on TikTok. Additionally, our dataset contained missing values (NA) for TikTok views, but it was unclear whether these represented songs with zero views or simply a lack of available data. Since we could not confidently differentiate between the two, we decided to remove all songs with missing TikTok data, which may have impacted our final dataset.

Second, our sampling method was not completely random. Songs were selected from Spotify playlists, meaning the dataset may not fully represent all popular songs within each genre. Some of these playlists may have included "discovery" tracks, which tend to feature less mainstream or emerging artists rather than only the most popular hits. As a result, the dataset may include some songs that are not widely known, while also potentially missing certain popular songs that gained traction through other means. This could introduce bias, particularly in terms of which songs were selected for analysis.

Despite these limitations, our study still offers valuable insights into the relationship between music genres and streaming platforms. The results can help advertisers optimize their targeting strategies and enable streaming services to refine their platform-specific approaches. Future research could improve upon this study by incorporating a broader selection of playlists or using alternative methods for genre classification to increase accuracy and representativeness.

Conclusion and Future Directions

By analyzing the role of release timing, platform-specific genre popularity, and the relationship between TikTok virality and Spotify streams, our project sheds light on the multifaceted ways digital platforms shape modern listening habits. We found that release month significantly affects Spotify streaming performance, with January standing out as a high-impact month—possibly due to reduced competition and renewed listener engagement. This insight can help artists and labels optimize when to launch new music to maximize exposure.

We also discovered that certain genres dominate specific platforms. For example, Pop performs especially well on TikTok, Country thrives on Pandora, and Latino music is most popular on YouTube. These patterns suggest that platform-user dynamics are genre-specific and that understanding these differences can help artists, record labels, and marketers tailor their content and advertising strategies more effectively.

Our third research question explored whether success on TikTok translates into Spotify streams. Despite some significant findings, like the strong predictive power of log_TikTok.Posts, our models revealed only a weak correlation. This highlights the limits of cross-platform influence, suggesting that while TikTok can help amplify a song's reach, it is not a guaranteed driver of Spotify's success. Other unmeasured factors likely play a major role in shaping what becomes a hit across different platforms.

These findings have important implications for both music creators and digital marketers. Artists and labels can use our insights to strategically plan release dates, choose which platforms to prioritize for promotion, and decide how to allocate resources across platforms. For example, partnering with influencers to build organic TikTok momentum might give a song a boost but, our results suggest this strategy should be combined with other promotional efforts to drive real traction on Spotify and other streaming platforms.

Streaming services like Spotify might also benefit from incorporating social media engagement data, like TikTok trends, into their recommendation systems. Doing so could help surface emerging songs in real time and align better with listener interest as it develops across platforms.
Ultimately, our study illustrates the interconnected but complex relationship between social media and music streaming. While TikTok, Spotify, and other platforms influence each other, their effects are not always linear or predictable. Future research can build on our work by incorporating additional platforms such as Instagram Reels or YouTube Shorts, using longitudinal data to track song performance over time, or integrating more detailed engagement metrics like TikTok shares, saves, or average watch time.

As social media continues to shape cultural trends and musical tastes, we hope our project contributes to a more data-driven understanding of how, when, and why songs go viral and what that means for the future of music.

References

Dawson, C. E., Jr., Mann, S., Roske, E., & Vasseur, G. (2021). Spotify: You have a Hit!. *SMU Data Science Review 5*(3), 9. https://scholar.smu.edu/datasciencereview/vol5/iss3/9

Jorgenson, L. (2022). *The influence of TikTok: Promotion trends in mainstream pop music*. https://scholarworks.calstate.edu/downloads/r781wp37g.pdf

Li, K. (2024). Predicting song popularity in the digital age through Spotify's data. *Theoretical and Natural Science*, *39*, 68-75. https://www.ewadirect.com/proceedings/tns/article/view/13412

Matos, B., Galuppo, F., Cordeiro, R., & Figueiredo, F. (2024). I've Heard This Before: Initial Results on Tiktok's Impact On the Re-Popularization of Songs. *arXiv preprint arXiv:2411.01239*. https://arxiv.org/pdf/2411.01239

MediaTech. (2023, September 20). *Music moves: Social media's influence on the modern music industry*. MediaTech Institute. https://mediatech.edu/music-moves-social-media-influence/

Middlebrook, K., & Sheik, K. (2019). Song hit prediction: Predicting billboard hits using Spotify data. *arXiv:1908.08609*. https://arxiv.org/pdf/1908.08609.pdf