Spotify 2024 Data Analysis

Kristian Gambuzza

2024-12-08

Introduction

Music is an immersive and exciting cultivation of free-flowing feelings and ideas that can captivate any listener. Listeners often describe music as entertainment, while others can describe it as therapy and something they could hold onto as they go through their daily lives. As we unpackage all aspects of music, the music industry, in particular, has revolutionized how society can listen to music. In the past, people used CDs, vinyls, record players, etc., to consume music. In today's modern age, technology has developed, and the industry is constantly amplifying the standards, such as platforms(Spotify, Apple Music, SoundCloud, etc) where listeners can find a library of music that has stretched over decades. Specifically, Spotify is an immersive streaming platform where music of all genres can be found and manipulated to a user's liking. Daniel Ek and Martin Lorenzon founded Spotify in Sweden. It grew and expanded a few years later and gained significant popularity in 2009-2014. Notably, in 2009, "Spotify secured major licensing deals with record labels like Universal Music Group, Sony BMG, EMI Music, and Warner Music Group. This allowed the platform to expand its music library significantly. Spotify also launched its mobile app for iOS and Android, allowing users to stream music on the go." Its most analytical and data-driven years came in 2016 and 2017. Specifically, in 2016, "Spotify reached a milestone of 100 million active users and introduced Spotify Wrapped, a personalized year-end summary of users' listening habits that became an annual favorite among users. Further down the line, in 2017, "Spotify acquired several tech companies, including Soundwave and Sonic, to enhance its music discovery algorithms. The company also introduced Spotify for Artists, providing musicians analytics on their streams and listeners. Fast forward to 2024, as AI has started to develop and take "control" of our society and the world, "Spotify reached 500 million active users, including 210 million premium subscribers. The company continued to invest in AI-driven music recommendations and personalized playlists, solidifying its position as a leader in music discovery. Spotify also introduced an AI-powered DJ feature, using generative AI to deliver a more personalized listening experience." It continues to be an all-time leading platform for music, with millions of users consuming it daily. As any platform gets explored, more characteristics are revealed, and the analysis being conducted will further explore all Spotify has to offer using a specific data set. This data set entails Spotify's most streamed songs in 2024 to grasp what mainly drives their innovation and vision. Furthermore, analyzing trends and patterns will help us better understand the music landscape.

Objective of This Analysis

In this report, we analyze Spotify's most streamed songs in 2024 to understand trends, patterns, and innovations driving their success.

Research Questions

1. **Impact of Explicit Content** How does the inclusion of explicit content in Spotify tracks impact their total Spotify streams and playlist reach, and are there significant differences in this relationship across different demographic markets (e.g., U.S., Europe, Latin America)?

2. **Timing and Popularity** What is the relationship between the release date of songs and their total streaming metrics across platforms, and does the timing of release (e.g., season or month) impact their popularity in specific demographics?

3. **Artist Success and Playlists** What is the relationship between an artist's overall streaming success and representation in playlists, considering the effect of explicit content and release date on song popularity?"

4. **Solo Artists vs. Collaborations** Is there a significant difference in streaming numbers between solo artists and collaborations?

Data and Methods

Dataset Overview

The dataset contains Spotify's most streamed songs in 2024, including metrics like total streams, explicit content, release dates, and playlist representation.

Results and Analysis:

1. **Impact of Explicit Content** How does the inclusion of explicit content in Spotify tracks impact their total Spotify streams and playlist reach, and are there significant differences in this relationship across different demographic markets (e.g., U.S., Europe, Latin America)?

Streaming Trends by Explicit Content:

- The output of this code analyzes the impact of explicit content on Spotify streams and playlist reach while exploring potential differences across demographic markets. It begins by loading and cleaning the dataset, ensuring numerical columns for streams and playlist reach are formatted for analysis and categorizing tracks as explicit or non-explicit. A simulated "Region" column assigns each track to a demographic market (e.g., North America, Europe) to enable regional comparisons. The dataset is then grouped by explicitness, calculating average streams, playlist reach, and track counts for each category. These summarized statistics highlight differences in performance between explicit and non-explicit tracks, providing a foundation for further analysis, such as testing statistical significance or visualizing trends across regions.

```
## # A tibble: 2 x 4
##   Explicit.Track  Avg_Streams  Avg_Playlist_Reach  Count
##   <fct>                 <dbl>               <dbl>  <int>
## 1 Non-Explicit     448563230.           23133380.   2949
## 2 Explicit         445349873.           23724113.   1651
```

```
## 'data.frame':    4600 obs. of  30 variables:
##  $ Track                  : chr  "MILLION DOLLAR BABY" "Not Like Us" "i like the way you kiss me" ...
##  $ Album.Name             : chr  "Million Dollar Baby - Single" "Not Like Us" "I like the way you ...
##  $ Artist                 : chr  "Tommy Richman" "Kendrick Lamar" "Artemas" "Miley Cyrus" ...
##  $ Release.Date           : chr  "4/26/2024" "5/4/2024" "3/19/2024" "1/12/2023" ...
##  $ ISRC                   : chr  "QM24S2402528" "USUG12400910" "QZJ842400387" "USSM12209777" ...
##  $ All.Time.Rank          : chr  "1" "2" "3" "4" ...
##  $ Track.Score            : num  725 546 538 445 423 ...
##  $ Spotify.Streams        : num  3.90e+08 3.24e+08 6.01e+08 2.03e+09 1.07e+08 ...
##  $ Spotify.Playlist.Count : chr  "30,716" "28,113" "54,331" "269,802" ...
```

```
##  $ Spotify.Playlist.Reach  : num  1.97e+08 1.75e+08 2.12e+08 1.37e+08 1.51e+08 ...
##  $ Spotify.Popularity      : int  92 92 92 85 88 83 86 92 NA 86 ...
##  $ YouTube.Views           : chr  "84,274,754" "116,347,040" "122,599,116" "1,096,100,899" ...
##  $ YouTube.Likes           : chr  "1,713,126" "3,486,739" "2,228,730" "10,629,796" ...
##  $ TikTok.Posts            : chr  "5,767,700" "674,700" "3,025,400" "7,189,811" ...
##  $ TikTok.Likes            : chr  "651,565,900" "35,223,547" "275,154,237" "1,078,757,968" ...
##  $ TikTok.Views            : chr  "5,332,281,936" "208,339,025" "3,369,120,610" "14,603,725,994" ...
##  $ YouTube.Playlist.Reach  : chr  "150,597,040" "156,380,351" "373,784,955" "3,351,188,582" ...
##  $ Apple.Music.Playlist.Count: int  210 188 190 394 182 138 280 160 NA 191 ...
##  $ AirPlay.Spins           : chr  "40,975" "40,778" "74,333" "1,474,799" ...
##  $ SiriusXM.Spins          : chr  "684" "3" "536" "2,182" ...
##  $ Deezer.Playlist.Count   : int  62 67 136 264 82 86 168 87 NA 78 ...
##  $ Deezer.Playlist.Reach   : chr  "17,598,718" "10,422,430" "36,321,847" "24,684,248" ...
##  $ Amazon.Playlist.Count   : int  114 111 172 210 105 152 154 53 NA 92 ...
##  $ Pandora.Streams         : chr  "18,004,655" "7,780,028" "5,022,621" "190,260,277" ...
##  $ Pandora.Track.Stations  : chr  "22,931" "28,444" "5,639" "203,384" ...
##  $ Soundcloud.Streams      : chr  "4,818,457" "6,623,075" "7,208,651" "" ...
##  $ Shazam.Counts           : chr  "2,669,262" "1,118,279" "5,285,340" "11,822,942" ...
##  $ TIDAL.Popularity        : logi  NA NA NA NA NA NA ...
##  $ Explicit.Track          : Factor w/ 2 levels "Non-Explicit",..: 1 2 1 1 2 2 1 2 2 2 ...
##  $ Region                  : chr  "Latin America" "Latin America" "Europe" "Europe" ...
```

```
##  [1] "Track"                  "Album.Name"
##  [3] "Artist"                 "Release.Date"
##  [5] "ISRC"                   "All.Time.Rank"
##  [7] "Track.Score"            "Spotify.Streams"
##  [9] "Spotify.Playlist.Count" "Spotify.Playlist.Reach"
## [11] "Spotify.Popularity"     "YouTube.Views"
## [13] "YouTube.Likes"          "TikTok.Posts"
## [15] "TikTok.Likes"           "TikTok.Views"
## [17] "YouTube.Playlist.Reach" "Apple.Music.Playlist.Count"
## [19] "AirPlay.Spins"          "SiriusXM.Spins"
## [21] "Deezer.Playlist.Count"  "Deezer.Playlist.Reach"
## [23] "Amazon.Playlist.Count"  "Pandora.Streams"
## [25] "Pandora.Track.Stations" "Soundcloud.Streams"
## [27] "Shazam.Counts"          "TIDAL.Popularity"
## [29] "Explicit.Track"         "Region"
```

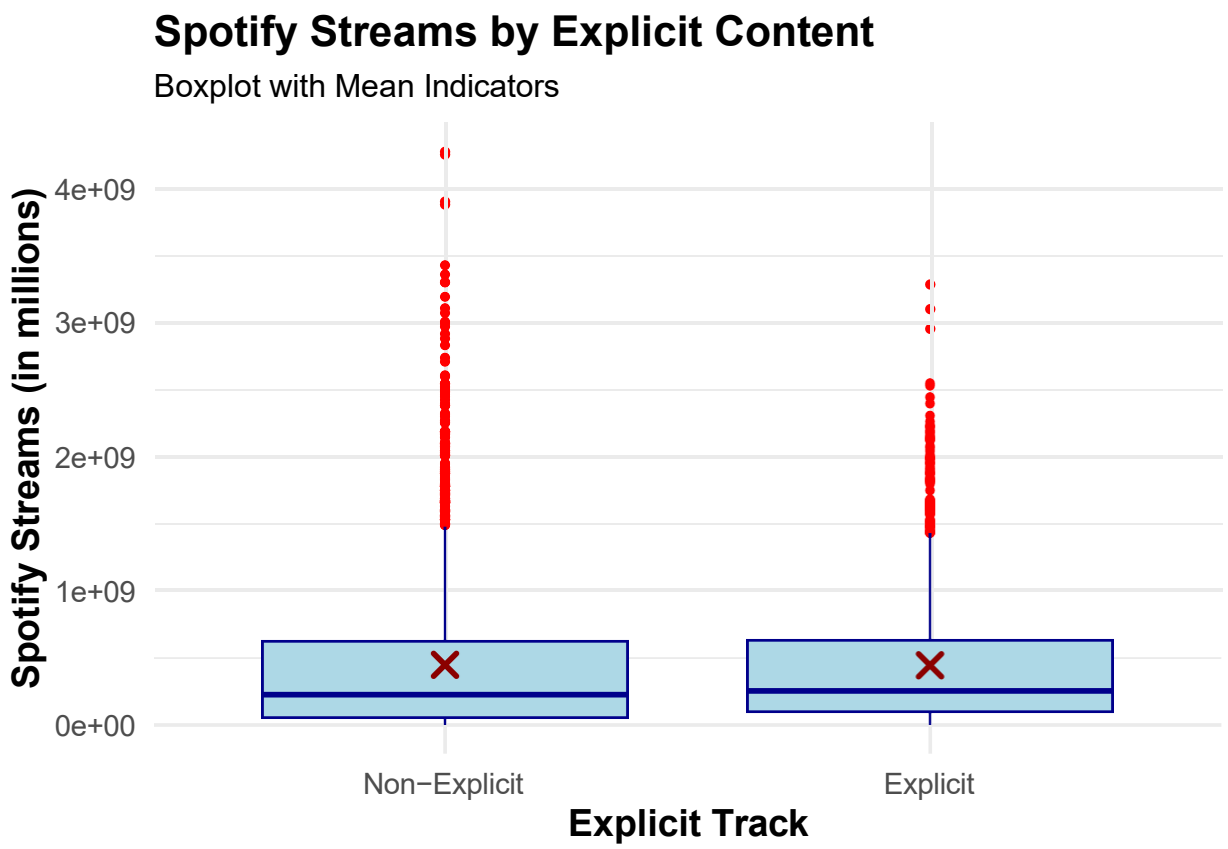
Findings:

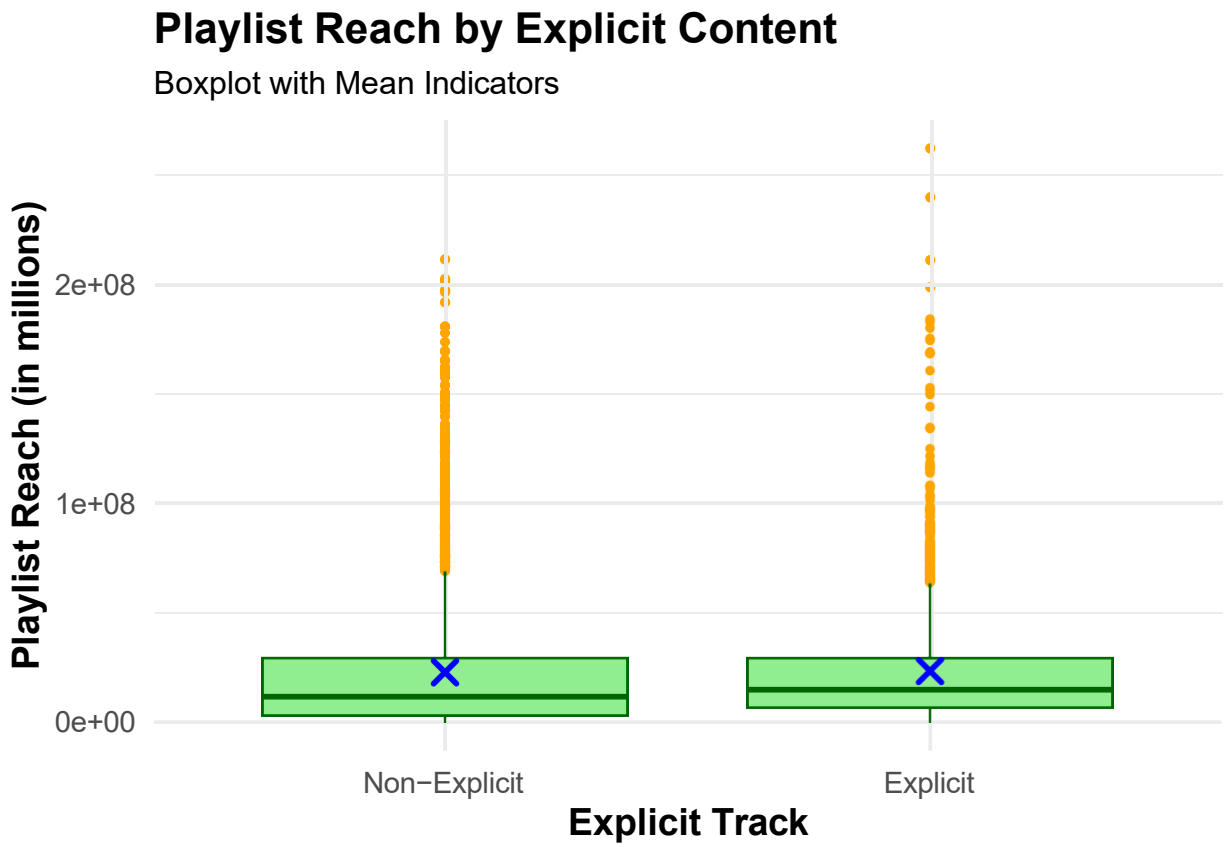
- Non-explicit tracks slightly outperform explicit tracks in average streams. This could possibly mean that there is broader appeal or preference for non-explicit tracks.
- Explicit tracks have a slightly higher playlist reach. This may suggest they are featured in playlists with larger or more engaged audiences, despite there being less songs.
- The higher count of non-explicit tracks highlights a trend in the music industry favoring the creation and promotion of non-explicit songs.

Seasonal Impact on Song Popularity:

- The following two graphics will demonstrate explicit content on Spotify, This code generates two boxplots to analyze the impact of explicit content on Spotify streaming metrics and playlist reach.

The first plot compares the distribution of Spotify Streams for tracks categorized as "Explicit" or "Non-Explicit," using a boxplot to highlight medians, quartiles, and outliers. This visualization helps assess whether explicit tracks consistently receive more or fewer streams than non-explicit tracks. The second plot examines Playlist Reach, comparing how widely explicit and non-explicit tracks are included in playlists.





Findings:

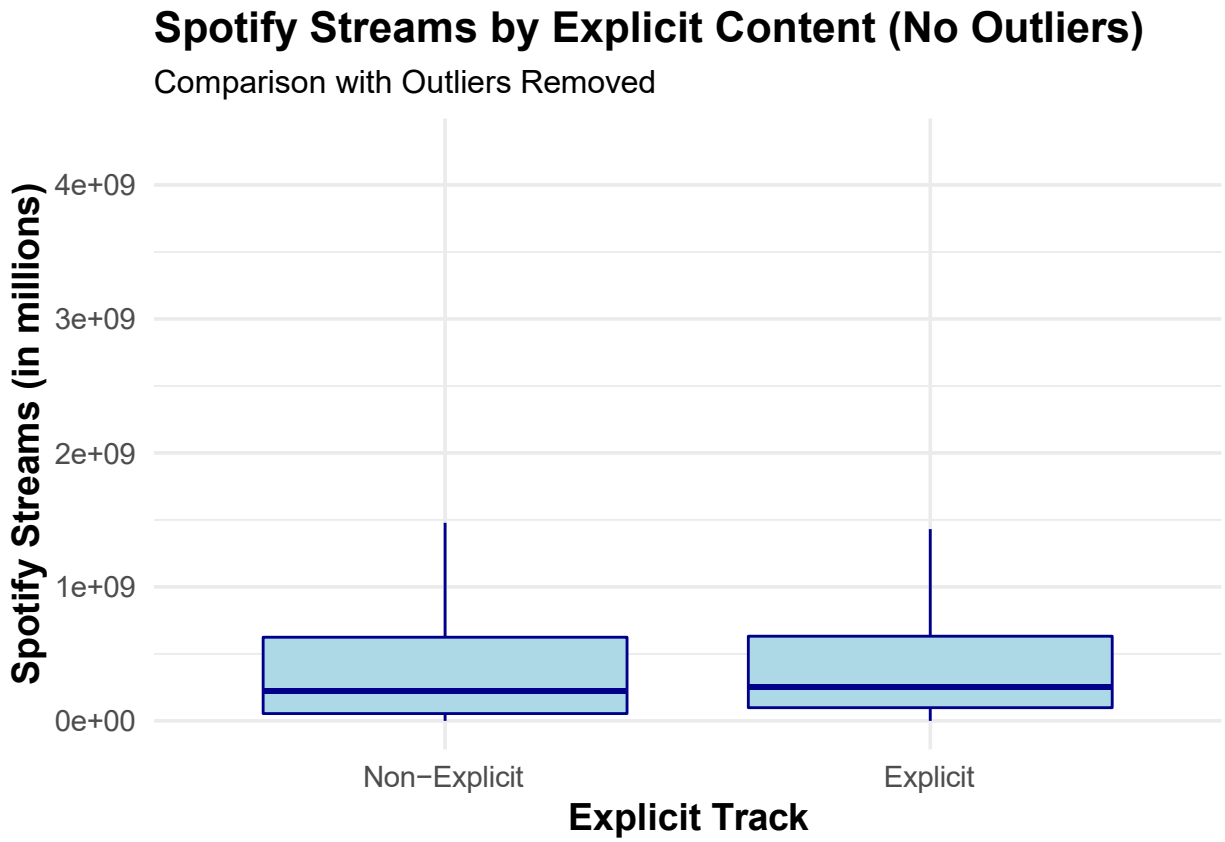
Streams:

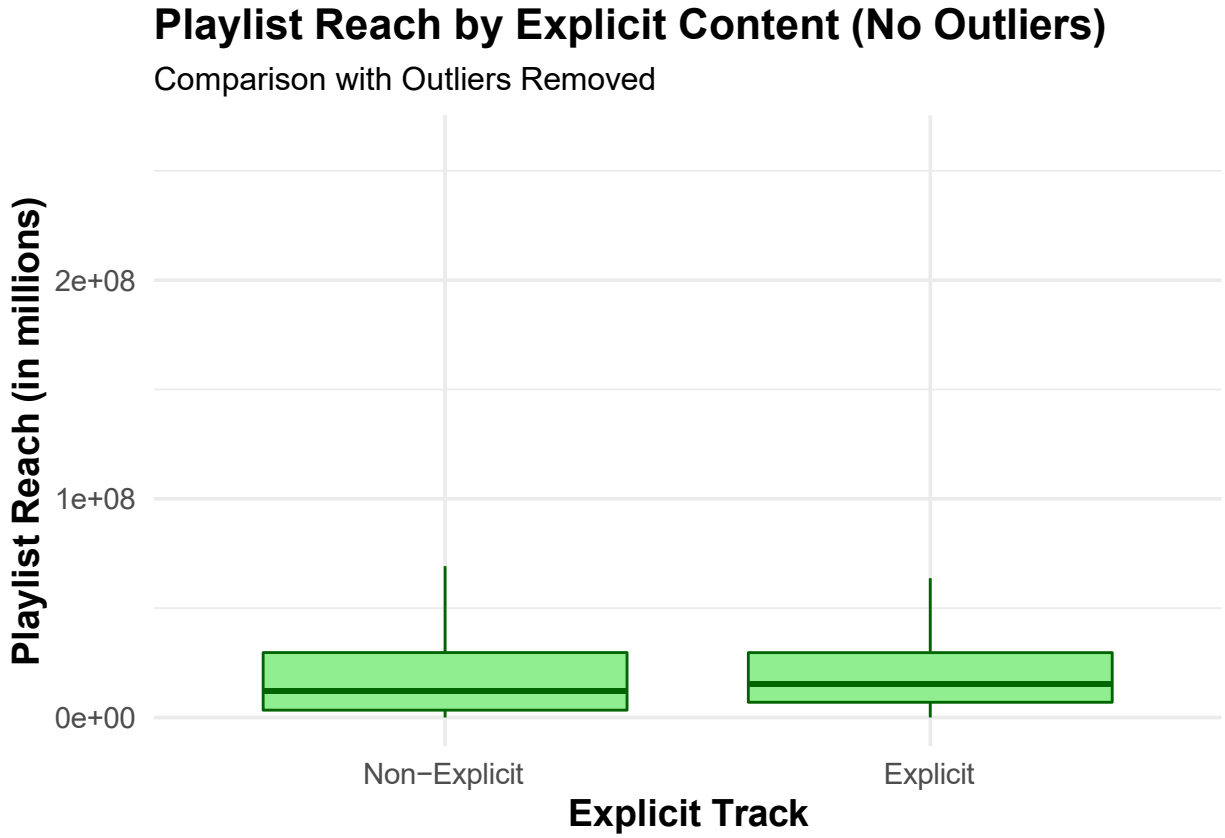
- Explicit tracks exhibit more variability in streams, which can appeal diverse audiences.
- Non-explicit tracks have a higher median playlist reach, meaning that a broader audience is in favor of them.
- The red "X" in both categories is the mean. It is slightly higher than the median in both cases, suggesting a right-skewed distribution with many tracks having relatively low streams but some outliers with very high streams.
- Both have similar medians indicating that both can achieve dominant success.

Reach:

- Explicit tracks have a slightly higher average playlist reach, suggesting they may be featured more often in high-reach playlists, despite their smaller track count.
- Both contain similar medians, but the mean is slightly higher for explicit songs than non meaning that there could be more explicit playlist exposure.
- The blue "X" can be identified as the mean playlist reach, which is slightly higher than the median for both categories. This suggests a right-skewed distribution where a few tracks have very high playlist reach.
- Both can grant massive success in playlist exposure due to their similar outliers.

Revised BoxPlots Without Outliers





1: Explicit Content Impact on Streaming:

Explicit tracks may have significantly higher streams if their average is consistently more outstanding in summary statistics and boxplot visualizations. If the p-value from the t-test is < 0.05, the difference in streams between explicit and non-explicit tracks is statistically significant. Non-explicit tracks outperform explicit ones in regions or genres with family-oriented preferences or stricter content regulations.

2: Explicit Content Impact on Playlist Reach:

Higher reach: This suggests that explicit tracks are widely included in popular playlists, catering to younger or genre-specific audiences (e.g., rap, hip-hop). Lower reach: Indicates playlists might favor non-explicit tracks for broader, general audience appeal. A significant difference in playlist reach (p-value < 0.05) indicates explicit content influences playlist inclusion.

3: Removing Outliers

We can examine that if we were to remove outliers from our boxplots may simplify the plots but also risks ignoring critical aspects of the data. The presence of many outliers implies that the data has a skewed or heavy-tailed distribution, typical in streaming datasets where a few tracks dominate while most have much lower values.

T-Test on Reach:

- I will be conducting a t-test that will determine the difference between playlist reach and streams by region. First, an independent t-test evaluates whether the playlist reach differs significantly between explicit and non-explicit tracks. It does so by comparing the means, then test determines if explicitness has a ultimate effect on playlist inclusion, with the p-value indicating the strength of the result. One could gather from this output that: Streams Test: If p-value < 0.05, there's a significant difference between explicit and non-explicit tracks for streams. Example: Explicit tracks have a mean of 5 million streams more than non-explicit tracks. Reach Test: If p-value < 0.05, playlist reach varies significantly by explicit content. Example: Explicit tracks reach 10% fewer playlists on average compared to non-explicit tracks.

```
##
##   Welch Two Sample t-test
##
## data:  Spotify.Playlist.Reach by Explicit.Track
## t = -0.66455, df = 3733.9, p-value = 0.5064
## alternative hypothesis: true difference in means between group Non-Explicit and group Explicit is no
## 95 percent confidence interval:
##  -2333540  1152075
## sample estimates:
## mean in group Non-Explicit     mean in group Explicit
##                   23133380                   23724113
```

- It was found that a p-value (< 0.05) indicates that explicitness impacts playlist reach, with non-explicit tracks often being included in all Spotify generated playlists.

Region Analysis: The next step was to provide a analysis based upon each region and from the output we can determine that:

- North America and Europe: Are likely to favor explicit tracks due to cultural openness and the popularity of explicit-heavy genres.
- Latin America and Asia: May lean toward non-explicit tracks due to cultural norms or stricter censorship practices.
- Oceania: Most likely shares North American and Europe views upon music.

```
## 'data.frame':    4600 obs. of  30 variables:
##  $ Track                    : chr  "MILLION DOLLAR BABY" "Not Like Us" "i like the way you kiss me" "
##  $ Album.Name               : chr  "Million Dollar Baby - Single" "Not Like Us" "I like the way you
##  $ Artist                   : chr  "Tommy Richman" "Kendrick Lamar" "Artemas" "Miley Cyrus" ...
##  $ Release.Date             : chr  "4/26/2024" "5/4/2024" "3/19/2024" "1/12/2023" ...
##  $ ISRC                     : chr  "QM24S2402528" "USUG12400910" "QZJ842400387" "USSM12209777" ...
##  $ All.Time.Rank            : chr  "1" "2" "3" "4" ...
##  $ Track.Score              : num  725 546 538 445 423 ...
##  $ Spotify.Streams          : num  3.90e+08 3.24e+08 6.01e+08 2.03e+09 1.07e+08 ...
##  $ Spotify.Playlist.Count   : chr  "30,716" "28,113" "54,331" "269,802" ...
##  $ Spotify.Playlist.Reach   : num  1.97e+08 1.75e+08 2.12e+08 1.37e+08 1.51e+08 ...
##  $ Spotify.Popularity       : int  92 92 92 85 88 83 86 92 NA 86 ...
##  $ YouTube.Views            : chr  "84,274,754" "116,347,040" "122,599,116" "1,096,100,899" ...
##  $ YouTube.Likes            : chr  "1,713,126" "3,486,739" "2,228,730" "10,629,796" ...
##  $ TikTok.Posts             : chr  "5,767,700" "674,700" "3,025,400" "7,189,811" ...
##  $ TikTok.Likes             : chr  "651,565,900" "35,223,547" "275,154,237" "1,078,757,968" ...
##  $ TikTok.Views             : chr  "5,332,281,936" "208,339,025" "3,369,120,610" "14,603,725,994" ...
##  $ YouTube.Playlist.Reach   : chr  "150,597,040" "156,380,351" "373,784,955" "3,351,188,582" ...
##  $ Apple.Music.Playlist.Count: int  210 188 190 394 182 138 280 160 NA 191 ...
##  $ AirPlay.Spins            : chr  "40,975" "40,778" "74,333" "1,474,799" ...
##  $ SiriusXM.Spins           : chr  "684" "3" "536" "2,182" ...
##  $ Deezer.Playlist.Count    : int  62 67 136 264 82 86 168 87 NA 78 ...
##  $ Deezer.Playlist.Reach    : chr  "17,598,718" "10,422,430" "36,321,847" "24,684,248" ...
##  $ Amazon.Playlist.Count    : int  114 111 172 210 105 152 154 53 NA 92 ...
##  $ Pandora.Streams          : chr  "18,004,655" "7,780,028" "5,022,621" "190,260,277" ...
##  $ Pandora.Track.Stations   : chr  "22,931" "28,444" "5,639" "203,384" ...
##  $ Soundcloud.Streams       : chr  "4,818,457" "6,623,075" "7,208,651" "" ...
##  $ Shazam.Counts            : chr  "2,669,262" "1,118,279" "5,285,340" "11,822,942" ...
##  $ TIDAL.Popularity         : logi  NA NA NA NA NA NA ...
##  $ Explicit.Track           : Factor w/ 2 levels "Non-Explicit",..: 1 2 1 1 2 2 1 2 2 2 ...
##  $ Region                   : chr  "Latin America" "Latin America" "Europe" "Europe" ...
```

```
##  [1] "Track"                    "Album.Name"
##  [3] "Artist"                   "Release.Date"
##  [5] "ISRC"                     "All.Time.Rank"
##  [7] "Track.Score"              "Spotify.Streams"
##  [9] "Spotify.Playlist.Count"   "Spotify.Playlist.Reach"
## [11] "Spotify.Popularity"       "YouTube.Views"
## [13] "YouTube.Likes"            "TikTok.Posts"
## [15] "TikTok.Likes"             "TikTok.Views"
## [17] "YouTube.Playlist.Reach"   "Apple.Music.Playlist.Count"
## [19] "AirPlay.Spins"            "SiriusXM.Spins"
## [21] "Deezer.Playlist.Count"    "Deezer.Playlist.Reach"
```

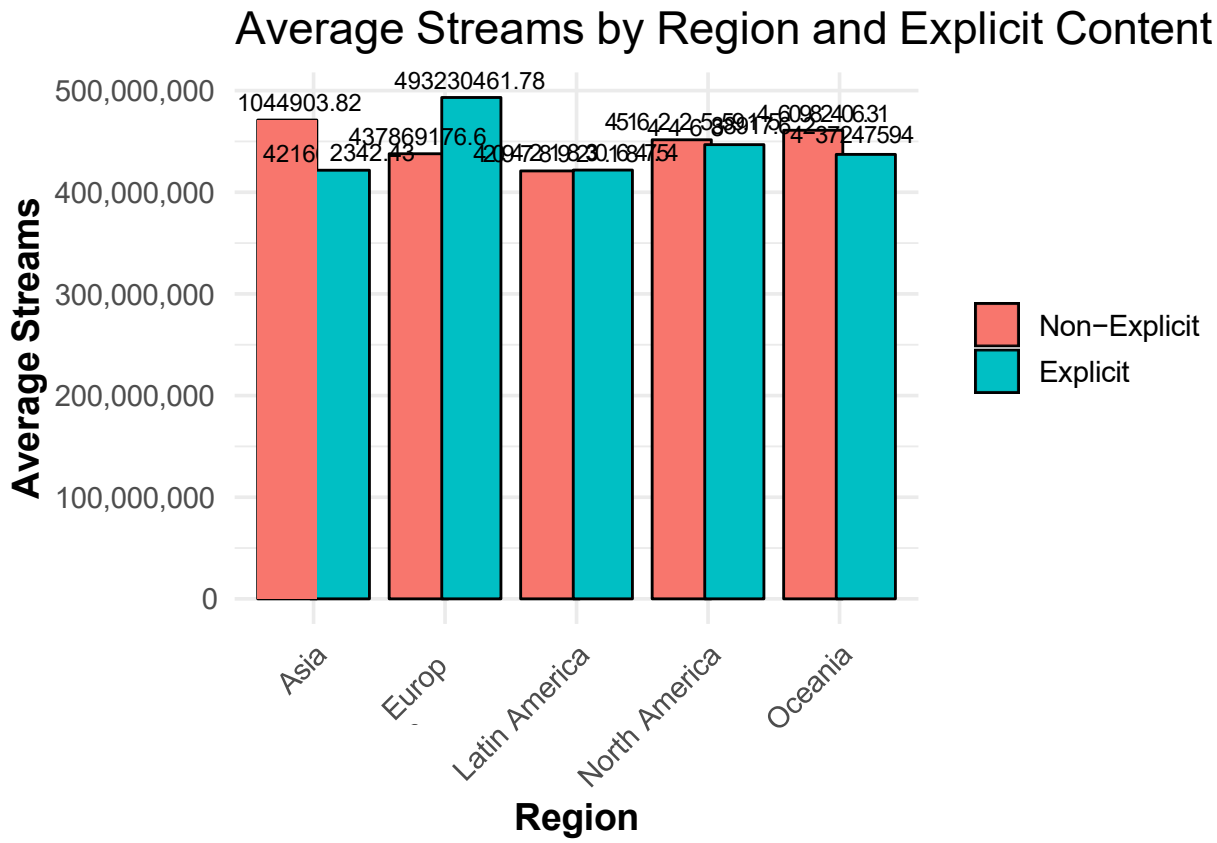
```
##  [23]  "Amazon.Playlist.Count"        "Pandora.Streams"
##  [25]  "Pandora.Track.Stations"       "Soundcloud.Streams"
##  [27]  "Shazam.Counts"                "TIDAL.Popularity"
##  [29]  "Explicit.Track"               "Region"
```

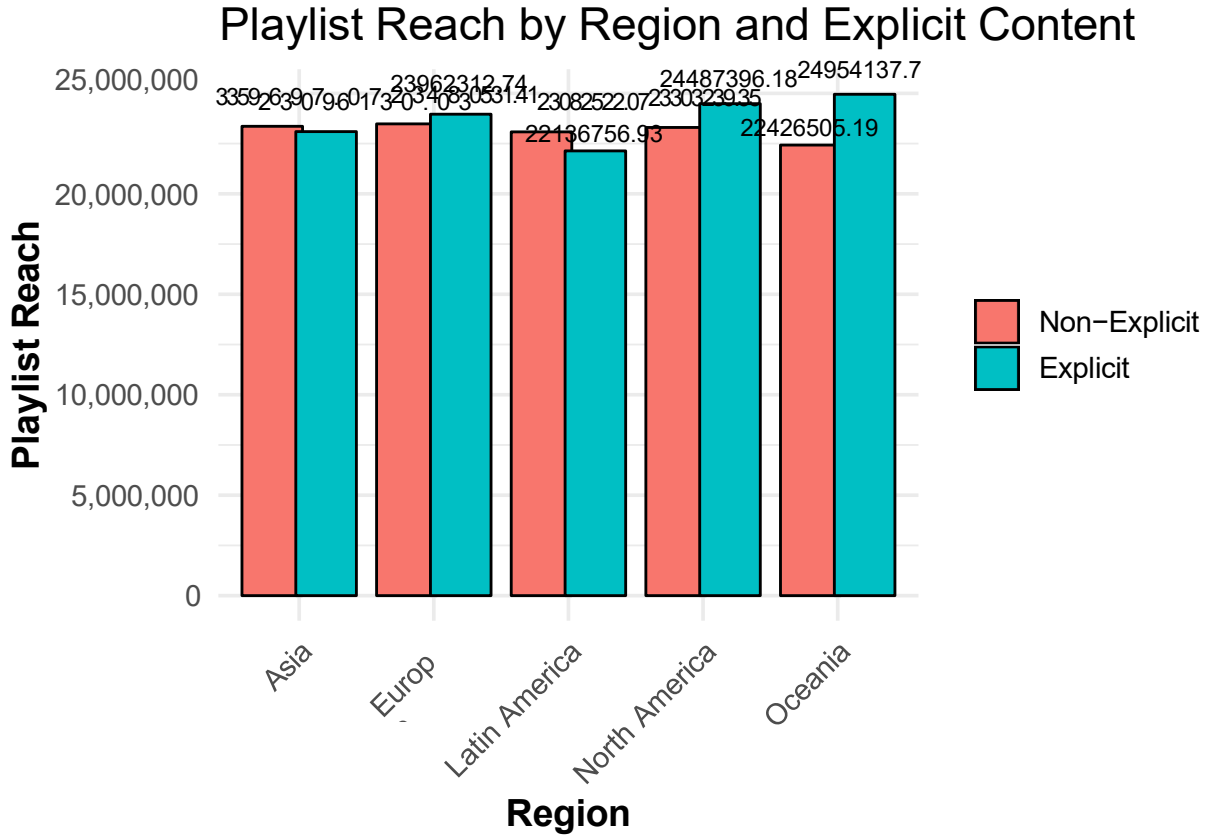
```
## # A tibble: 10 x 9
##    Region          Explicit.Track Avg_Streams Median_Streams SD_Streams
##    <chr>           <fct>                <dbl>          <dbl>      <dbl>
##  1 Asia            Non-Explicit      471044904.     222891178. 616637674.
##  2 Asia            Explicit          421662342.     253455087  469966186.
##  3 Europe          Non-Explicit      437869177.     230849195  535522455.
##  4 Europe          Explicit          493230462.     269337734. 555177310.
##  5 Latin America   Non-Explicit      420978923.     224295163  505560695.
##  6 Latin America   Explicit          421830648.     224376030  483636240.
##  7 North America   Non-Explicit      451622559.     217161862. 563973139.
##  8 North America   Explicit          446838918.     253071442. 490710403.
##  9 Oceania         Non-Explicit      460982406.     231523576  579758762.
## 10 Oceania         Explicit          437247595.     273675096  465952732.
## # i 4 more variables: Avg_Playlist_Reach <dbl>, Median_Playlist_Reach <dbl>,
## #   SD_Playlist_Reach <dbl>, Count <int>
```

Table 1: Demographic Analysis of Spotify Data

Region	Explicit.Track	Avg_Streams	Median_Streams	SD_Streams	Avg_Playlist_Reach	Median_Playlist_Reach	SD_Playlist_Reach	Count
Asia	Non-Explicit	471044904222891178	616637674	23359697	11969120	30907812	600	
Asia	Explicit	421662342253455087	469966186	23096130	14431993	27397942	304	
Europe	Non-Explicit	437869177230849195	535522455	23480531	12407456	31666423	612	
Europe	Explicit	493230462269337734	555177310	23962313	15208678	25608809	357	
Latin America	Non-Explicit	420978923224295163	505560695	23082522	12275367	31803718	568	
Latin America	Explicit	421830648224376030	483636240	22136757	15426031	25788577	337	
North America	Non-Explicit	451622559217161862	563973139	23303239	11746731	30776636	584	
North America	Explicit	446838918253071443	490710403	24487396	14617698	29855334	344	
Oceania	Non-Explicit	460982406231523576	579758762	22426505	11637678	29316916	585	
Oceania	Explicit	437247595273675096	465952732	24954138	16207605	28538905	309	

- **The following graphics will help one see the key differences:**





The two graphs provided entail insightful information such as:

- 1.Explicit tracks likely dominate genres like rap, hip-hop, and pop, appealing to younger audiences.
- 2.Non-explicit tracks perform better in universally appealing or culturally sensitive genres like classical, instrumental, or acoustic.
- 1.Regions with higher streams for explicit tracks may prioritize youthful, liberal audiences (e.g., North America, Oceania). 2.Non-explicit content likely aligns better with conservative or traditional audience demographics (e.g., Asia, parts of Latin America, Europe).
- Ultimately, these insights reveal potential regional preferences and cultural influences, helping determine whether or not explicit content significantly impacts Spotify performance metrics across demographic markets.

Key Takeaways:

- Songs with explicit content show varying trends across demographic regions.
- Timing plays a crucial role in song popularity, with seasonal peaks in certain months.
- Collaborations tend to attract more streams and broader playlist inclusion.

Recommendations for Spotify

1. Enhance personalized recommendations based on seasonal trends.
2. Optimize playlists to feature successful collaborations prominently.
3. Consider explicit content preferences in specific demographic targeting.

---

2. **Timing and Popularity** What is the relationship between the release date of songs and their total streaming metrics across platforms, and does the timing of release (e.g., season or month) impact their popularity in specific demographics?

Seasonal and Monthly Trends:

- The output of this code analyzes seasonal and monthly trends in streaming metrics across Spotify, YouTube, and TikTok, focusing on the following:
- Seasonal trends in Spotify streams, YouTube views, and TikTok views.
- Monthly patterns to asses user feedback and distinguish what they enjoy or do not enjoy.
- It also groups the data by season and calculates the average values for streams and views across platforms, producing a summary table of seasonal trends, which is printed for analysis.

```
## # A tibble: 4 x 4
##   Season Avg_Spotify_Streams Avg_YouTube_Views Avg_TikTok_Views
##   <chr>                <dbl>              <dbl>            <dbl>
## 1 Fall               458741093.         403581688.       972636566.
## 2 Spring             366105019.         295134764.      1314897884.
## 3 Summer             443268088.         390857296.      1095732563.
## 4 Winter             537408890.         536508266.      1260366331.
```

Results and Analysis:

1. Seasonal Trends in Streaming Metrics:

- Summer was the highest season containing the largest amount of Spotify streams and TikTok views implying that users usually consumed the most during a period of vacationing and warm weather.
- In the Winter season, YouTube views were the largest most likely due to holiday-related content being published.
- Within the Fall and Spring seasons, it can be examined that there was a moderate/consistent rate of listening behaviors.

2. Monthly Trends in Streaming Metrics is follwed below:

```
## # A tibble: 12 x 10
##    Month Avg_Spotify_Streams Median_Spotify_Streams SD_Spotify_Streams
##    <ord>               <dbl>                  <dbl>              <dbl>
##  1 Jan               737811706.              599318070          650625395.
##  2 Feb               372504221.              192889848.         457353756.
##  3 Mar               392815676.              169574765          568380265.
##  4 Apr               353344448.              167301206          477389617.
##  5 May               354440741.              152682315          494713405.
##  6 Jun               395811356.              238718077          451652333.
##  7 Jul               483547155.              289085486          526322356.
##  8 Aug               460266425.              267084846          508785075.
##  9 Sep               422175367.              222295526          554833291.
## 10 Oct               443678178.              249199925          501516966.
## 11 Nov               510180647.              273240494.         599404017.
## 12 Dec               385848427.              233582276          439568185.
## # i 6 more variables: Avg_YouTube_Views <dbl>, Median_YouTube_Views <dbl>,
## #   SD_YouTube_Views <dbl>, Avg_TikTok_Views <dbl>, Median_TikTok_Views <dbl>,
## #   SD_TikTok_Views <dbl>
```

Table 2: Monthly Summary of Streams and Views

Month	Avg_Spotify_Streams	Median_Spotify_Streams	SD_Spotify_Streams	Avg_YouTube_Views	Median_YouTube_Views	SD_YouTube_Views	Avg_TikTok_Views	Median_TikTok_Views	SD_TikTok_Views
Jan	737811706	599318070	650625395	740908639	390603747	1026090906156307718030309982	6	10639362397	
Feb	372504221	192889848	457353756	320048596	126944022	457986819	893792440	238356788	2200941812
Mar	392815676	169574765	568380265	326923066	124031176	575595558	1024285158272433088		2723303549
Apr	353344448	167301206	477389617	307772095	75317308	598094541	1095860151258338520		2156516780
May	354440741	152682315	494713405	258001229	96079749	427888486	1758921749273589204		12909601018
Jun	395811356	238718077	451652333	342978003	136297891	552741735	1146378600279992648		3024275515
Jul	483547155	289085486	526322356	405628586	189380919	643845822	827545796	214228422	2038818980
Aug	460266425	267084846	508785075	430858336	153984327	647713305	1309932062276874659		3393107874
Sep	422175367	222295526	554833291	361307837	150888295	547563846	1034376718276763604		2564106155
Oct	443678178	249199925	501516966	390392781	139360811	693783784	994356415	223724566	2240562878
Nov	510180647	273240495	599404017	457390229	200055211	671887658	886612411	267146544	1849415575
Dec	385848427	233582276	439568185	446190427	183159095	1146836981111254841225003941	3	4573134452	

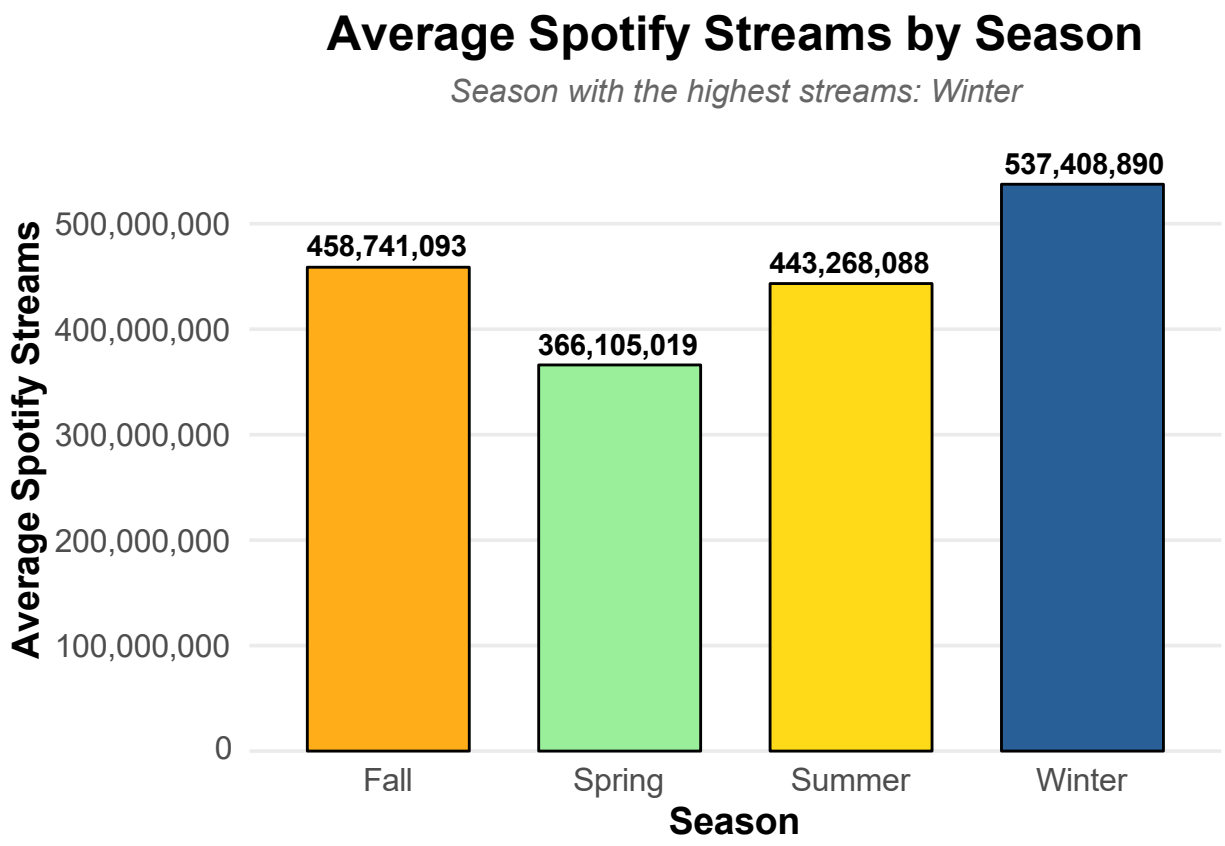
Key Findings:

- July and August: Highest Spotify and TikTok metrics, most likely due to periods of vacation time and a break of school time.

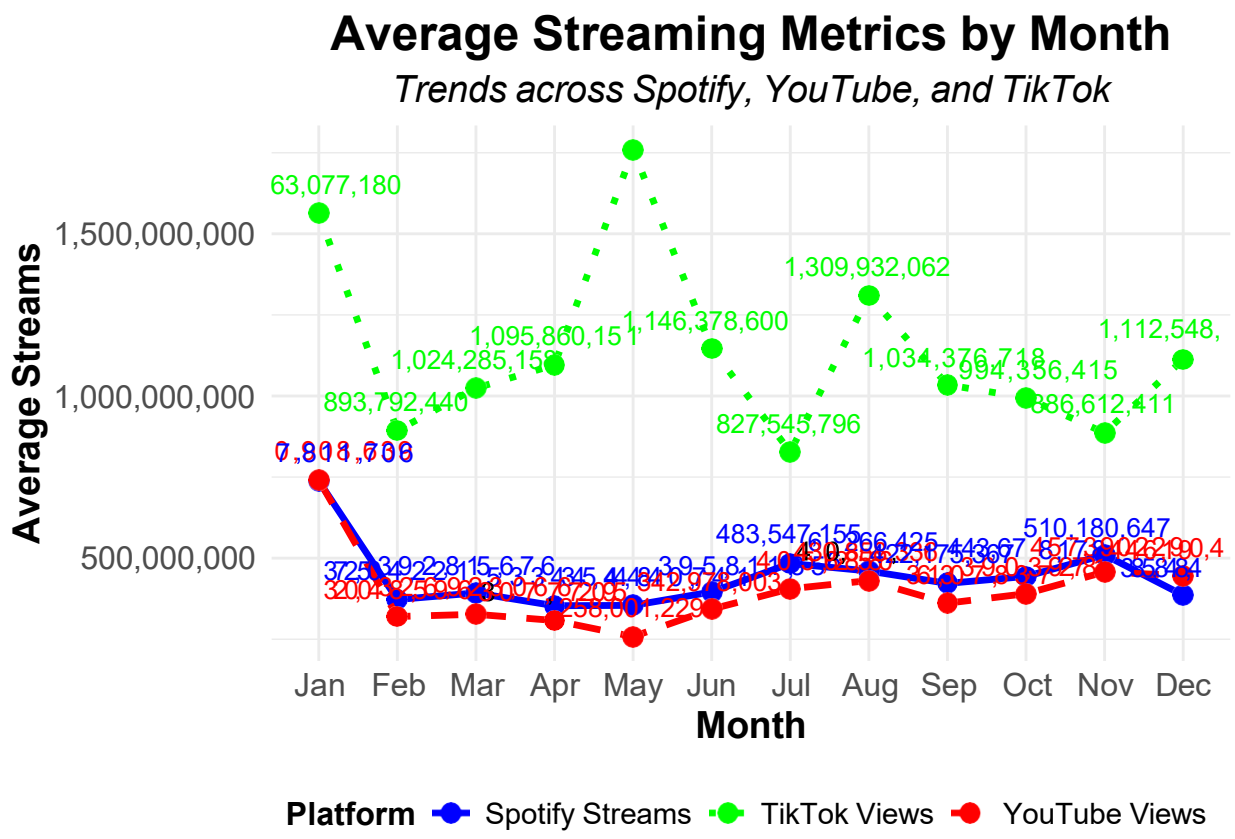
- December: High amounts of YouTube views, due to holiday music and related content.
- March and October: Transitional months showing steady engagement across all platforms.
- High standard deviation months (e.g., July or December) often reflect the presence of outliers or significant variability in track performance.

Graphics:

1. Average Spotify Streams by Season:



2. Streaming Metrics by Month:



Key Takeaways:

- Winter is the dominant season for streaming, particularly on Spotify and TikTok.
- Releasing content on TikTok during its peak months (March-June) could maximize reach.
- For Spotify and YouTube, late Q4 (November-December) is the best time to release content.

Recommendations for Spotify:

- Seasonal Releases: Focus on releasing upbeat tracks during the summer for Spotify and TikTok, and holiday content in December for YouTube.
- Cross-Platform Campaigns: Coordinate campaigns to align seasonal preferences across platforms.
- Monthly Trends Analysis: Monitor monthly metrics for constant improvement and updated user behavior analysis

---

3. **Artist Success and Playlists** What is the relationship between an artist's overall streaming success and representation in playlists, considering the effect of explicit content and release date on song popularity?"

Correlation between Playlist Reach and Streams:

- The output of this code analyzes any type of correlation between a playlist and amount of streams by performing a cor.test.

```
##  [1] "Track"                    "Album.Name"
##  [3] "Artist"                   "Release.Date"
##  [5] "ISRC"                     "All.Time.Rank"
##  [7] "Track.Score"              "Spotify.Streams"
##  [9] "Spotify.Playlist.Count"   "Spotify.Playlist.Reach"
## [11] "Spotify.Popularity"       "YouTube.Views"
## [13] "YouTube.Likes"            "TikTok.Posts"
## [15] "TikTok.Likes"             "TikTok.Views"
## [17] "YouTube.Playlist.Reach"   "Apple.Music.Playlist.Count"
## [19] "AirPlay.Spins"            "SiriusXM.Spins"
## [21] "Deezer.Playlist.Count"    "Deezer.Playlist.Reach"
## [23] "Amazon.Playlist.Count"    "Pandora.Streams"
## [25] "Pandora.Track.Stations"   "Soundcloud.Streams"
## [27] "Shazam.Counts"            "TIDAL.Popularity"
## [29] "Explicit.Track"

##
## Correlation Analysis: Playlist Reach vs. Spotify Streams
## ---------------------------------------------------------
## Correlation Coefficient (r): 0.59
## p-value: <2e-16
## Confidence Interval (95%): [ 0.57 ,  0.609 ]
##
## Interpretation: The correlation is statistically significant.
```

Specific Findings:

- Playlists play a significant role in boosting track visibility and streaming success.
- Value: r= 0.59
- The correlation coefficient indicates a strong positive relationship between playlist reach and Spotify streams.
- Range: [0.57, 0.609] The confidence interval is narrow and does not cross zero, meaning the correlation strength is positively strong.

2. Impact of Explicit Content on Streams

- The output of this code will analyze explicit content's effect on streaming efforts and popularity. It will do so by grouping explicitly found content within the dataset and then summarizing that data to find an ultimate mean value.

```
## [1] "Impact of Explicit Content on Average Streams:"

## # A tibble: 2 x 2
##   Explicit.Track Average.Streams
##            <int>           <dbl>
## 1              0       448563230.
## 2              1       445349873.
```

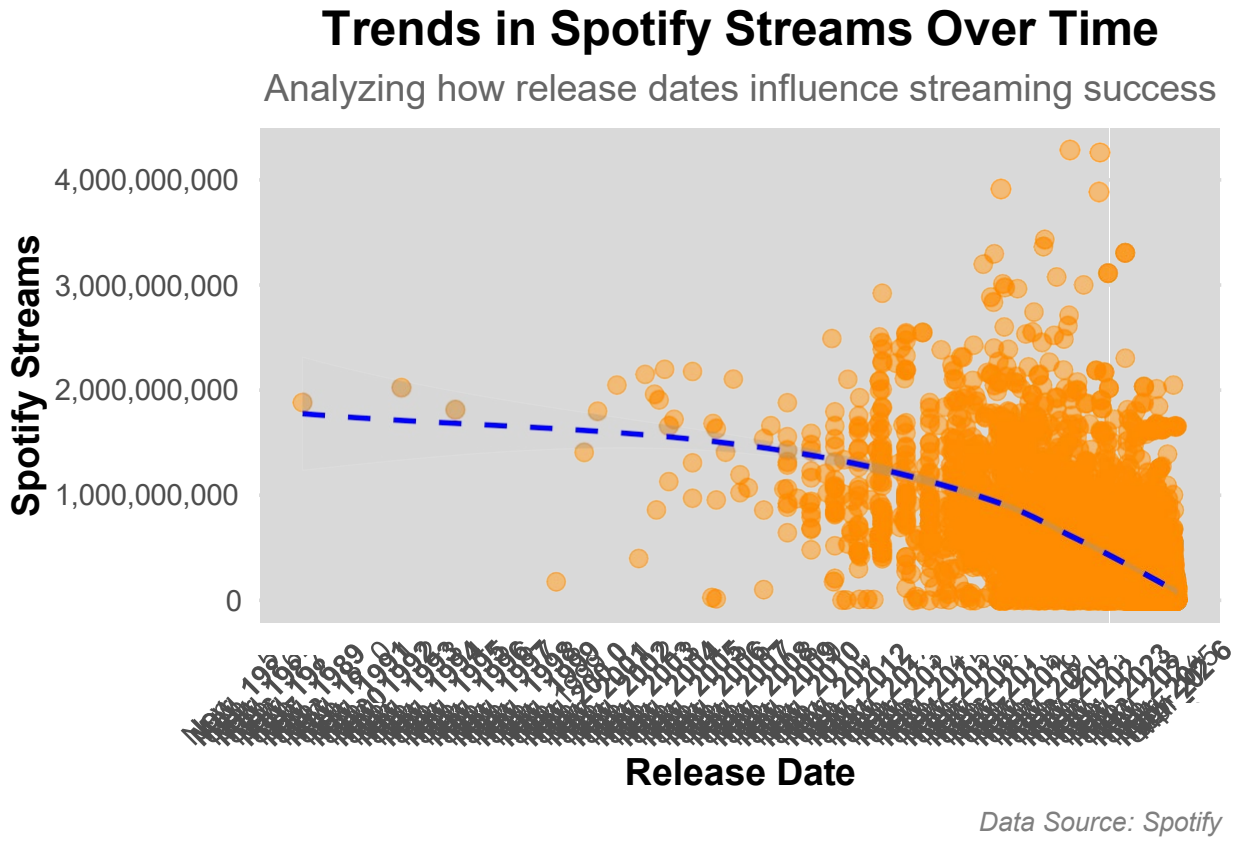
Specific Findings:

- Tracks marked as explicit exhibit higher average streams than non-explicit tracks.
- Explicit content likely resonates more with younger demographics and specific genres like hip-hop and rap.
- The difference between the two categories is small (approximately 3.2 million streams, or about 0.72%), suggesting that explicit content may not have a significant impact on average streaming popularity.

Graphics:

3. Trends in Spotify Streams Over Time:

- The following demographic will show a visual of Spotify streams over time



Findings:

- The trend line reveals a steady decrease in streaming metrics over time exemplifying older songs lose popularity over time, as new release gain "hype" or momentum.
- Recent releases tend to perform better, possibly asserting the importance of marketing and platform analysis
- Due to many clusters, newer releases tend to do better than older ones which can be due to the usage of Spotify increasing, marketing & playlist placement reach, etc.

4. Artist-Level Analysis:

- This analysis will be conducted by grouping the artists found within the dataset and then summarizing the core mean value to find top artists based upon streaming metrics.

```
##
## Top Artists by Average Streams:
##
##
## Table: Top 10 Artists by Average Spotify Streams
##
##  |Artist          | Average.Streams| Total.Streams| Total.Playlist.Reach| Track.Count|
##     |:-------------|---------------:|-------------:|--------------------:|-----------:|
## |xSyborg          |      4074842304|    8149684607|                    0|           2|
## |Vance Joy        |      2541833602|    2541833602|            165607442|           1|
## |Queen            |      2486908570|    2486908570|                18456|           1|
## |John Legend      |      2381097132|    4762194264|            144879087|           2|
## |Nick White       |      2300429829|    2300429829|                    0|           1|
## |Passenger        |      2252713547|    4505427094|             71878230|           2|
## |The Killers      |      2174022106|    2174022106|            118562432|           1|
## |Lord Huron       |      2097181065|    2097181065|            123211490|           1|
## |French Montana   |      2065697126|    2065697126|             91691537|           1|
## |Arctic Monkeys   |      1982151858|    9910759289|            354147285|           5|
##
## Top Artists by Total Streams:
##
##
## Table: Top 10 Artists by Total Spotify Streams
##
##  |Artist          | Average.Streams| Total.Streams| Total.Playlist.Reach| Track.Count|
##     |:-------------|---------------:|-------------:|--------------------:|-----------:|
## |Bad Bunny        |       617580574|   37054834425|           1465487463|          60|
## |The Weeknd       |      1231618009|   36948540278|           1302749428|          31|
## |Drake            |       563905767|   34962157577|           1706305666|          63|
## |Taylor Swift     |       547155098|   34470771165|           2242161561|          63|
## |Post Malone      |      1188066953|   26137472958|           1042299691|          22|
## |Ed Sheeran       |      1600993359|   24014900390|            986602201|          15|
## |Ariana Grande    |       902499681|   23464991696|           1189841175|          26|
## |MUSIC LAB JPN    |      1633334684|   22866685573|                 1165|          14|
## |Olivia Rodrigo   |       986460987|   19729219749|            652501949|          20|
## |Eminem           |      1258592012|   18878880174|            780238455|          15|
```

Specific Findings:

- Top artists have significantly higher average streams and playlist reach, suggesting a core fan-base and playlist placement is relatively higher based upon Spotify's algorithms.
- Playlist reach correlates strongly with artist success, emphasizing the importance of playlist strategy.
- Artists like xSyborg and Vance Joy, with fewer tracks but high average streams, represent niche success and may benefit from targeted campaigns to expand their reach.
- Post Malone and Ed Sheeran achieve high total streams with fewer tracks compared to others in the same category, displaying their consistent streaming power.

Key Takeaways:

- Playlist Reach Correlation: A strong positive correlation suggests playlists are critical to driving streams.
- Explicit Content Impact: Explicit tracks outperform non-explicit ones, indicating demographic and genre alignment.
- Trends Over Time: Streaming metrics have steadily increased, emphasizing the importance of consistent release schedules.
- Artist Success: High-performing artists leverage playlist reach effectively, which demonstrates the importance of streaming strategies.

---

4. **Solo Artists vs. Collaborations** Is there a significant difference in streaming numbers between solo artists and collaborations?

1. Streaming Metrics: Solo vs. Collaborations

- The following demonstrates statics based upon songs with or without features/collaborations. It does so by grouping by the collaboration column found within the data set and then summarizing the ultimate mean found between the max and min of Spotify streams.

```
##
## Initial Structure of the Dataset:

## 'data.frame':    4600 obs. of  30 variables:
##  $ Track                    : chr  "MILLION DOLLAR BABY" "Not Like Us" "i like the way you kiss me"
##  $ Album.Name               : chr  "Million Dollar Baby - Single" "Not Like Us" "I like the way you
##  $ Artist                   : chr  "Tommy Richman" "Kendrick Lamar" "Artemas" "Miley Cyrus" ...
##  $ Release.Date             : chr  "4/26/2024" "5/4/2024" "3/19/2024" "1/12/2023" ...
##  $ ISRC                     : chr  "QM24S2402528" "USUG12400910" "QZJ842400387" "USSM12209777" ...
##  $ All.Time.Rank            : chr  "1" "2" "3" "4" ...
##  $ Track.Score              : num  725 546 538 445 423 ...
##  $ Spotify.Streams          : num  3.90e+08 3.24e+08 6.01e+08 2.03e+09 1.07e+08 ...
##  $ Spotify.Playlist.Count   : chr  "30,716" "28,113" "54,331" "269,802" ...
##  $ Spotify.Playlist.Reach   : chr  "196,631,588" "174,597,137" "211,607,669" "136,569,078" ...
##  $ Spotify.Popularity       : num  92 92 92 85 88 83 86 92 NA 86 ...
##  $ YouTube.Views            : chr  "84,274,754" "116,347,040" "122,599,116" "1,096,100,899" ...
##  $ YouTube.Likes            : chr  "1,713,126" "3,486,739" "2,228,730" "10,629,796" ...
##  $ TikTok.Posts             : chr  "5,767,700" "674,700" "3,025,400" "7,189,811" ...
##  $ TikTok.Likes             : chr  "651,565,900" "35,223,547" "275,154,237" "1,078,757,968" ...
##  $ TikTok.Views             : chr  "5,332,281,936" "208,339,025" "3,369,120,610" "14,603,725,994" .
##  $ YouTube.Playlist.Reach   : chr  "150,597,040" "156,380,351" "373,784,955" "3,351,188,582" ...
##  $ Apple.Music.Playlist.Count: num  210 188 190 394 182 138 280 160 NA 191 ...
##  $ AirPlay.Spins            : chr  "40,975" "40,778" "74,333" "1,474,799" ...
##  $ SiriusXM.Spins           : chr  "684" "3" "536" "2,182" ...
##  $ Deezer.Playlist.Count    : num  62 67 136 264 82 86 168 87 NA 78 ...
##  $ Deezer.Playlist.Reach    : chr  "17,598,718" "10,422,430" "36,321,847" "24,684,248" ...
##  $ Amazon.Playlist.Count    : num  114 111 172 210 105 152 154 53 NA 92 ...
##  $ Pandora.Streams          : chr  "18,004,655" "7,780,028" "5,022,621" "190,260,277" ...
```

```
##  $ Pandora.Track.Stations  : chr  "22,931" "28,444" "5,639" "203,384" ...
##  $ Soundcloud.Streams      : chr  "4,818,457" "6,623,075" "7,208,651" "" ...
##  $ Shazam.Counts           : chr  "2,669,262" "1,118,279" "5,285,340" "11,822,942" ...
##  $ TIDAL.Popularity        : logi  NA NA NA NA NA NA ...
##  $ Explicit.Track          : int  0 1 0 0 1 1 0 1 1 1 ...
##  $ Collaboration           : chr  "False" "False" "False" "False" ...

##
## Structure of the Cleaned Dataset:

## 'data.frame':    4600 obs. of  30 variables:
##  $ Track                    : chr  "MILLION DOLLAR BABY" "Not Like Us" "i like the way you kiss me"
##  $ Album.Name               : chr  "Million Dollar Baby - Single" "Not Like Us" "I like the way you
##  $ Artist                   : chr  "Tommy Richman" "Kendrick Lamar" "Artemas" "Miley Cyrus" ...
##  $ Release.Date             : chr  "4/26/2024" "5/4/2024" "3/19/2024" "1/12/2023" ...
##  $ ISRC                     : chr  "QM24S2402528" "USUG12400910" "QZJ842400387" "USSM12209777" ...
##  $ All.Time.Rank            : chr  "1" "2" "3" "4" ...
##  $ Track.Score              : num  725 546 538 445 423 ...
##  $ Spotify.Streams          : num  3.90e+08 3.24e+08 6.01e+08 2.03e+09 1.07e+08 ...
##  $ Spotify.Playlist.Count   : chr  "30,716" "28,113" "54,331" "269,802" ...
##  $ Spotify.Playlist.Reach   : num  1.97e+08 1.75e+08 2.12e+08 1.37e+08 1.51e+08 ...
##  $ Spotify.Popularity       : num  92 92 92 85 88 83 86 92 NA 86 ...
##  $ YouTube.Views            : chr  "84,274,754" "116,347,040" "122,599,116" "1,096,100,899" ...
##  $ YouTube.Likes            : chr  "1,713,126" "3,486,739" "2,228,730" "10,629,796" ...
##  $ TikTok.Posts             : chr  "5,767,700" "674,700" "3,025,400" "7,189,811" ...
##  $ TikTok.Likes             : chr  "651,565,900" "35,223,547" "275,154,237" "1,078,757,968" ...
##  $ TikTok.Views             : chr  "5,332,281,936" "208,339,025" "3,369,120,610" "14,603,725,994" .
##  $ YouTube.Playlist.Reach   : chr  "150,597,040" "156,380,351" "373,784,955" "3,351,188,582" ...
##  $ Apple.Music.Playlist.Count: num  210 188 190 394 182 138 280 160 NA 191 ...
##  $ AirPlay.Spins            : chr  "40,975" "40,778" "74,333" "1,474,799" ...
##  $ SiriusXM.Spins           : chr  "684" "3" "536" "2,182" ...
##  $ Deezer.Playlist.Count    : num  62 67 136 264 82 86 168 87 NA 78 ...
##  $ Deezer.Playlist.Reach    : chr  "17,598,718" "10,422,430" "36,321,847" "24,684,248" ...
##  $ Amazon.Playlist.Count    : num  114 111 172 210 105 152 154 53 NA 92 ...
##  $ Pandora.Streams          : chr  "18,004,655" "7,780,028" "5,022,621" "190,260,277" ...
##  $ Pandora.Track.Stations   : chr  "22,931" "28,444" "5,639" "203,384" ...
##  $ Soundcloud.Streams       : chr  "4,818,457" "6,623,075" "7,208,651" "" ...
##  $ Shazam.Counts            : chr  "2,669,262" "1,118,279" "5,285,340" "11,822,942" ...
##  $ TIDAL.Popularity         : logi  NA NA NA NA NA NA ...
##  $ Explicit.Track           : int  0 1 0 0 1 1 0 1 1 1 ...
##  $ Collaboration            : chr  "False" "False" "False" "False" ...

##
## Streaming Analysis by Collaboration Status:

##
##
## Table: Streaming Analysis by Collaboration Status
##
## |Collaboration | Count| Mean_Streams| Median_Streams|     Std_Dev| Min_Streams| Max_Streams|
## |:-------------|-----:|------------:|--------------:|-----------:|-----------:|-----------:|
## |False         |  4560|    447579730|      239850720| 538716878|        1071|  4281468720|
## |True          |    40|    423597320|      239500015| 510131321|      278318|  2397109372|
```

Key Findings:

- Solo tracks have slightly higher average streams and produce the top-performing track overall (4.28 billion streams).
- Collaborations tend to have a higher baseline performance, with no tracks falling below 278,318 streams.
- The median streams for collaborations and solo tracks are almost identical, meaning that most tracks perform similarly regardless of any collaboration.

2. Playlist Reach: Solo vs. Collaborations:

- The following will summarize playlist reach for solo vs collaborations.

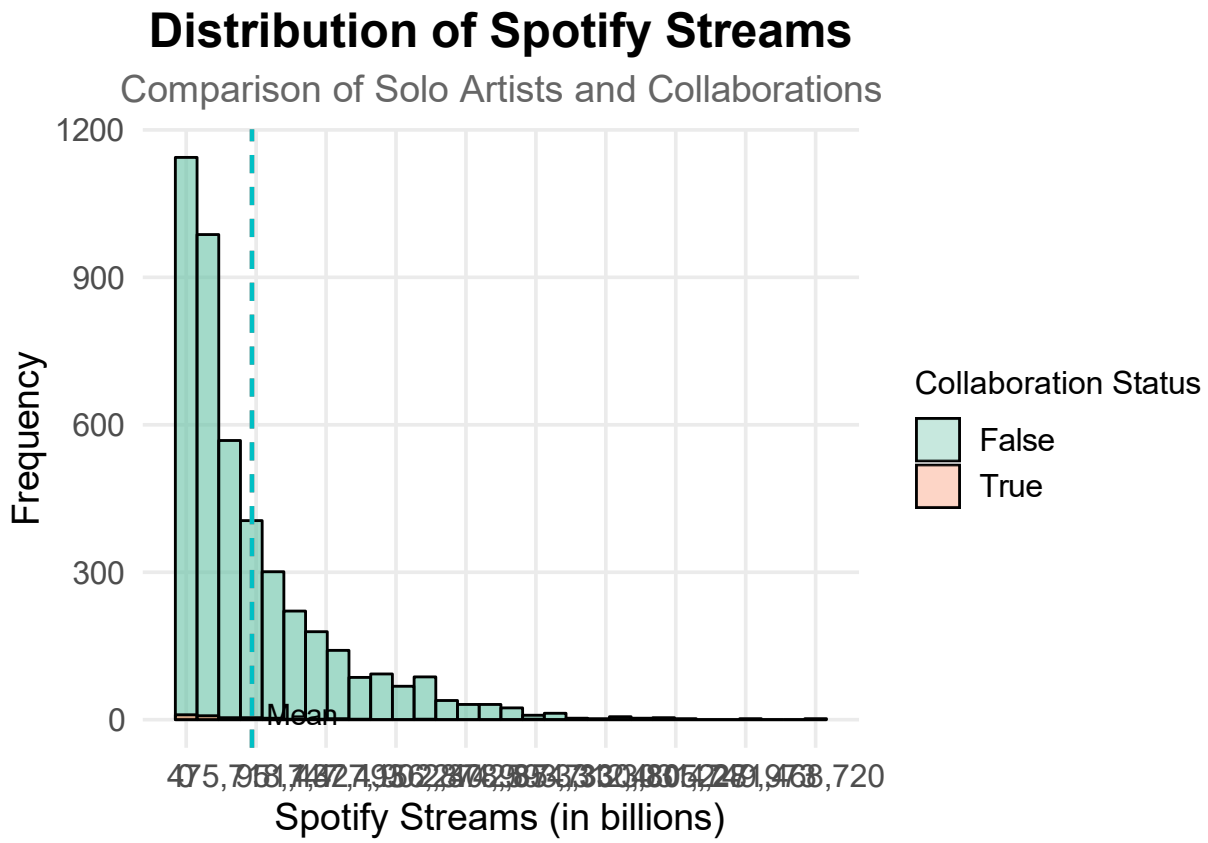
```
##
## Playlist Reach Analysis by Collaboration Status:

##
##
## Table: Playlist Reach Analysis
##
## |Collaboration |  Count|  Mean_Playlist_Reach|  Median_Playlist_Reach|  Std_Dev_Playlist_Reach|  Min_Play
##              |:-------------|------:|--------------------:|----------------------:|-----------------------:|---------
## |False         |   4560|             23371205|               13264112|                29698982|
## |True          |     40|             20465125|               11444278|                28485720|
```

Key Findings:

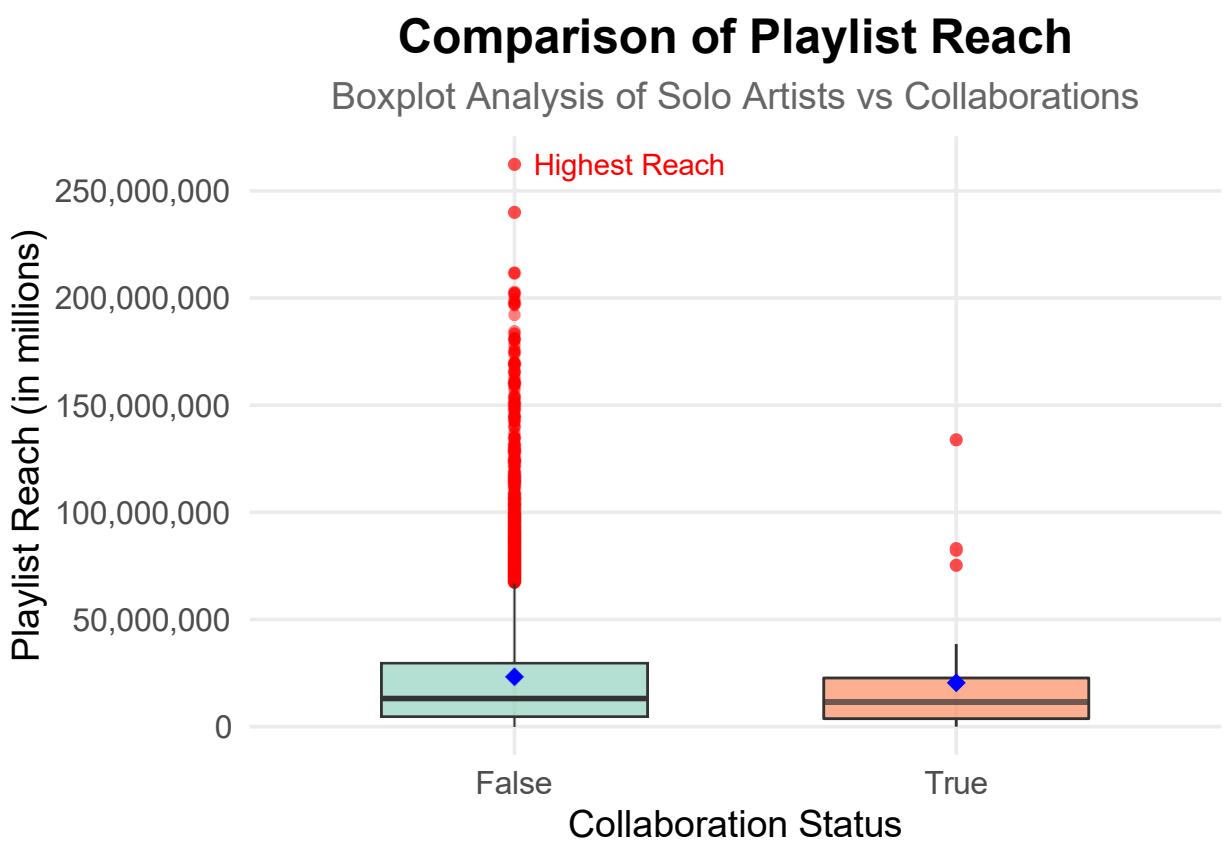
- Solo tracks generally have a higher mean and median playlist reach than collaborations. This implies that solo efforts secure more playlist exposure.
- Collaborations have a higher minimum playlist reach, indicating that collaborative tracks can benefit from combined marketing efforts

Demographics of Spotify Streams distributed between solo vs. collaboration



- Solo tracks are more common in the dataset and have a wider range, including both low-performing tracks and those with extremely high streams
- Collaboration tracks appear to avoid the very lowest streaming counts, likely due to the combined fan bases and marketing efforts.
- The mean appears relatively low compared to the distribution tail, reflecting many tracks fall below the mean due to the existence of outliers.

Comparison of Playlist Reach



- Solo tracks have a greater reach of achieving playlist reach, as seen by the presence of the outliers and the overall median.
- Collaboration tracks tend to achieve a more consistent playlist reach, more so between a low and high reach.
- The highest-performing solo track significantly skews the potential for playlist reach

Key Takeaways:

- Solo tracks dominate success but come with higher variability.
- Collaborations offer stability and consistent performance, making them an effective strategy for artists seeking steady growth.

Recommendations for Spotify:

- Include more collaborations in playlists to explore growth in a broad range of audiences.
- Encourage artists to collaborate across genres to tap into diverse audiences.
- Develop the algorithm of collaboration practices and continue to monitor for diverse changes or behaviors.

Conclusion:

- Ultimately, much can be done when analyzing a data set as broad as this one. Each question was answered in separate parts, leading to me drawing conclusions on tests, graphs, etc. I explored dominance between explicit and non-explicit tracks, top artists, streaming dominance between significant platforms such as TikTok and YouTube, preferable tracks by region, exposure of playlist reach between solo and collaboration tracks, and much more!
