

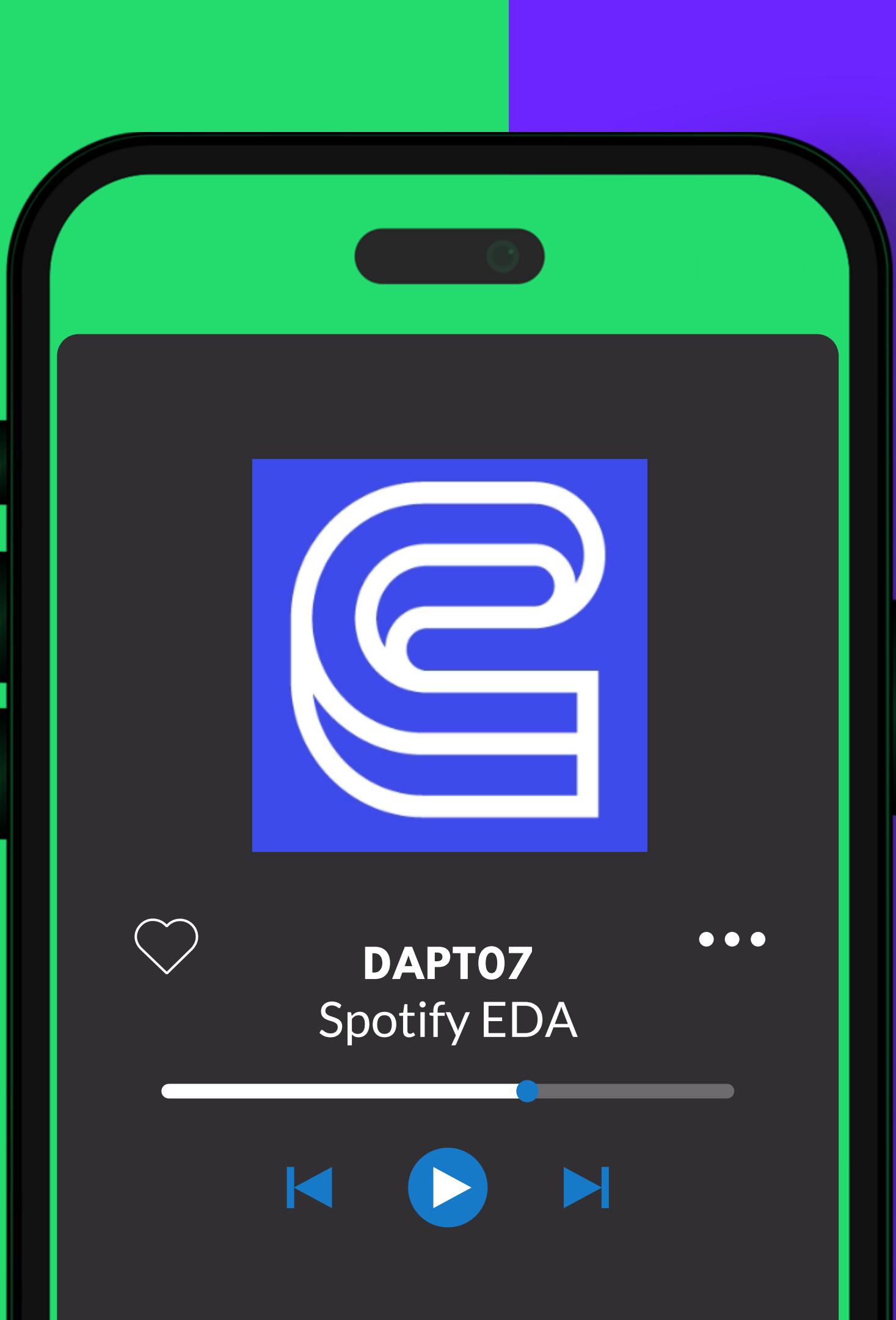
Business Insights From Spotify

PRESENTATION

Karina Edwards
Yeng Russian

Agenda

- **Data Presentation**
 - **Introduction to the topic**
 - **Description of Dataset (data overview)**
- **Data Pre-processing**
- **Data Processing**
 - **Manipulate data and answer Questions**
- **Key Insights**
- **Conclusion**
- **Comments and Limitations**





Spotify®

2006

Swedish audio streaming and media service provider

640 Mio+

users

100 Mio+

tracks



Data Overview

Most Relevant Variables

Track

Name or title of the song.

Artist

Name of the performer(s) or band that recorded the track.

Release Date

When the track was officially released.

Track Score

Aggregate score that represents the overall performance of a track.



Data Overview

Most Relevant Variables

Spotify Streams

Total number of times the track has been streamed.

Spotify Popularity

Current popularity of the track

YouTube Views/Likes

Total number of views /likes per track in Youtube.

TikTok Views/Likes

Total number of views /likes per track in TikTok.



Quick Overview

	Track	Album Name	Artist	Release Date	ISRC	All Time Rank	Track Score	Spotify Streams	Spotify Playlist Count	Spotify Playlist Reach	...	SiriusXM Spins	Deezer Playlist Count	Deezer Playlist Reach	Amazon Playlist Count	Pandora Streams	Pandora Track Stations
0	MILLION DOLLAR BABY - Single	Million Dollar Baby - Single	Tommy Richman	4/26/2024	QM24S2402528	1	725.4	390,470,936	30,716	196,631,588	...	684	62.0	17,598,718	114.0	18,004,655	22
1	Not Like Us	Not Like Us	Kendrick Lamar	5/4/2024	USUG12400910	2	545.9	323,703,884	28,113	174,597,137	...	3	67.0	10,422,430	111.0	7,780,028	28
2	i like the way you kiss me	I like the way you kiss me	Artemas	3/19/2024	QZJ842400387	3	538.4	601,309,283	54,331	211,607,669	...	536	136.0	36,321,847	172.0	5,022,621	5
3	Flowers	Flowers - Single	Miley Cyrus	1/12/2023	USSM12209777	4	444.9	2,031,280,633	269,802	136,569,078	...	2,182	264.0	24,684,248	210.0	190,260,277	203
4	Houdini	Houdini	Eminem	5/31/2024	USUG12403398	5	423.3	107,034,922	7,223	151,469,874	...	1	82.0	17,660,624	105.0	4,493,884	7

4600
rows

- Origin of the dataset: Based on the US data.
- 'Release Date' variable is in US format.
- It shall not be considered as a Worldwide metric.

data.info()				
<class 'pandas.core.frame.DataFrame'>				
RangeIndex: 4600 entries, 0 to 4599				
Data columns (total 29 columns):				
#	Column	Non-Null Count	Dtype	
0	Track	4600	non-null	object
1	Album Name	4600	non-null	object
2	Artist	4595	non-null	object
3	Release Date	4600	non-null	object
4	ISRC	4600	non-null	object
5	All Time Rank	4600	non-null	object
6	Track Score	4600	non-null	float64
7	Spotify Streams	4487	non-null	object
8	Spotify Playlist Count	4530	non-null	object
9	Spotify Playlist Reach	4528	non-null	object
10	Spotify Popularity	3796	non-null	float64
11	YouTube Views	4292	non-null	object
12	YouTube Likes	4285	non-null	object
13	TikTok Posts	3427	non-null	object
14	TikTok Likes	3620	non-null	object
15	TikTok Views	3619	non-null	object
16	YouTube Playlist Reach	3591	non-null	object
17	Apple Music Playlist Count	4039	non-null	float64
18	AirPlay Spins	4182	non-null	object
19	SiriusXM Spins	2477	non-null	object
20	Deezer Playlist Count	3679	non-null	float64
21	Deezer Playlist Reach	3672	non-null	object
22	Amazon Playlist Count	3545	non-null	float64
23	Pandora Streams	3494	non-null	object
24	Pandora Track Stations	3332	non-null	object
25	Soundcloud Streams	1267	non-null	object
26	Shazam Counts	4023	non-null	object
27	TIDAL Popularity	0	non-null	float64
28	Explicit Track	4600	non-null	int64
dtypes: float64(6), int64(1), object(22)				
memory usage: 1.0+ MB				

Data Pre-processing

2 duplicated rows

```
data.drop_duplicates(keep='first', inplace=True)  
data.shape
```

110 duplicated tracks

```
data = data.drop_duplicates(subset=['Track', 'Artist'], keep='first')  
data.shape
```

5 datatype changes

```
data_cleaned['Spotify Streams'] = data_cleaned['Spotify Streams'].str.replace(',', '').astype(float)  
  
data_cleaned['All Time Rank'] = data_cleaned['All Time Rank'].str.replace(',', '').astype(float)  
  
data_cleaned['YouTube Likes'] = data_cleaned['YouTube Likes'].str.replace(',', '').astype(float)  
  
data_cleaned['TikTok Likes'] = data_cleaned['TikTok Likes'].str.replace(',', '').astype(float)  
  
data_cleaned['Spotify Playlist Reach'] = data_cleaned['Spotify Playlist Reach'].str.replace(',', '').astype(float)
```

NaN values

```
data_cleaned = data.dropna(subset=['Artist'])  
  
data_cleaned = data_cleaned.dropna(subset=['Spotify Streams'])  
  
data_cleaned = data_cleaned.dropna(subset=['YouTube Likes'])  
  
data_cleaned = data_cleaned.dropna(subset=['TikTok Likes'])  
  
data_cleaned = data_cleaned.dropna(subset=['Spotify Popularity'])
```

Data Pre-processing

Special Characters removal

```
data_cleaned = data_cleaned[~data_cleaned['Artist'].str.contains('X', na = False)]
data_cleaned = data_cleaned[~data_cleaned['Artist'].str.contains('Y', na = False)]
data_cleaned = data_cleaned[~data_cleaned['Artist'].str.contains('í', na = False)]
print(data_cleaned)
```

Data index reset

```
data_cleaned.reset_index(drop=True, inplace = True)
data_cleaned.info()
```

Data Sampling

```
data_cleaned.sample(n=10)
```

	Track	Album Name	Artist	Release Date	ISRC	All Time Rank	Track Score	Spotify Streams	Spotify Playlist Count	Spotify Playlist Reach	...	Pandora Streams	Pandora Track Stations	Soundcloud Streams	Shazam Counts
243	Adderall (Corvette Corvette)	Adderall (Corvette Corvette)	Popp Hunna	2020-10-14	USA2P2036916	332.0	84.1	90383224.0	24,868	4716437.0	...	6,283,417	26,512	10,556,085	759,001
1417	After Party	Heaven Or Hell	Don Toliver	2020-03-13	USAT22001223	2005.0	32.2	706711575.0	170,468	34103511.0	...	11,521,920	46,060	NaN	1,400,011
35	If We Ever Broke Up	If We Ever Broke Up	Mae Stephens	2023-02-10	GBUM72300969	49.0	203.6	304929492.0	43,923	46133083.0	...	7,444,354	7,188	NaN	2,227,494
2793	Hit Different	Hit Different	SZA	2020-09-04	USRC12002998	4097.0	21.0	254979267.0	50,550	13267378.0	...	64,371,686	18,949	1,578,488	961,364
3002	085 - Ao Vivo	Pode Criar (Ao V	MC Rogerinho	2023-01-12	US7VG2308218	4401.0	19.9	126063862.0	10,150	9712919.0	...	NaN	NaN	NaN	97,522



Get ready for
The Analysis

What are we

01 Overall Distribution of popularity

02 Top Performers

03 Changes in Ratings

04 User Engagement

05 Trends across Platforms

Overall Popularity Across songs

Outliers

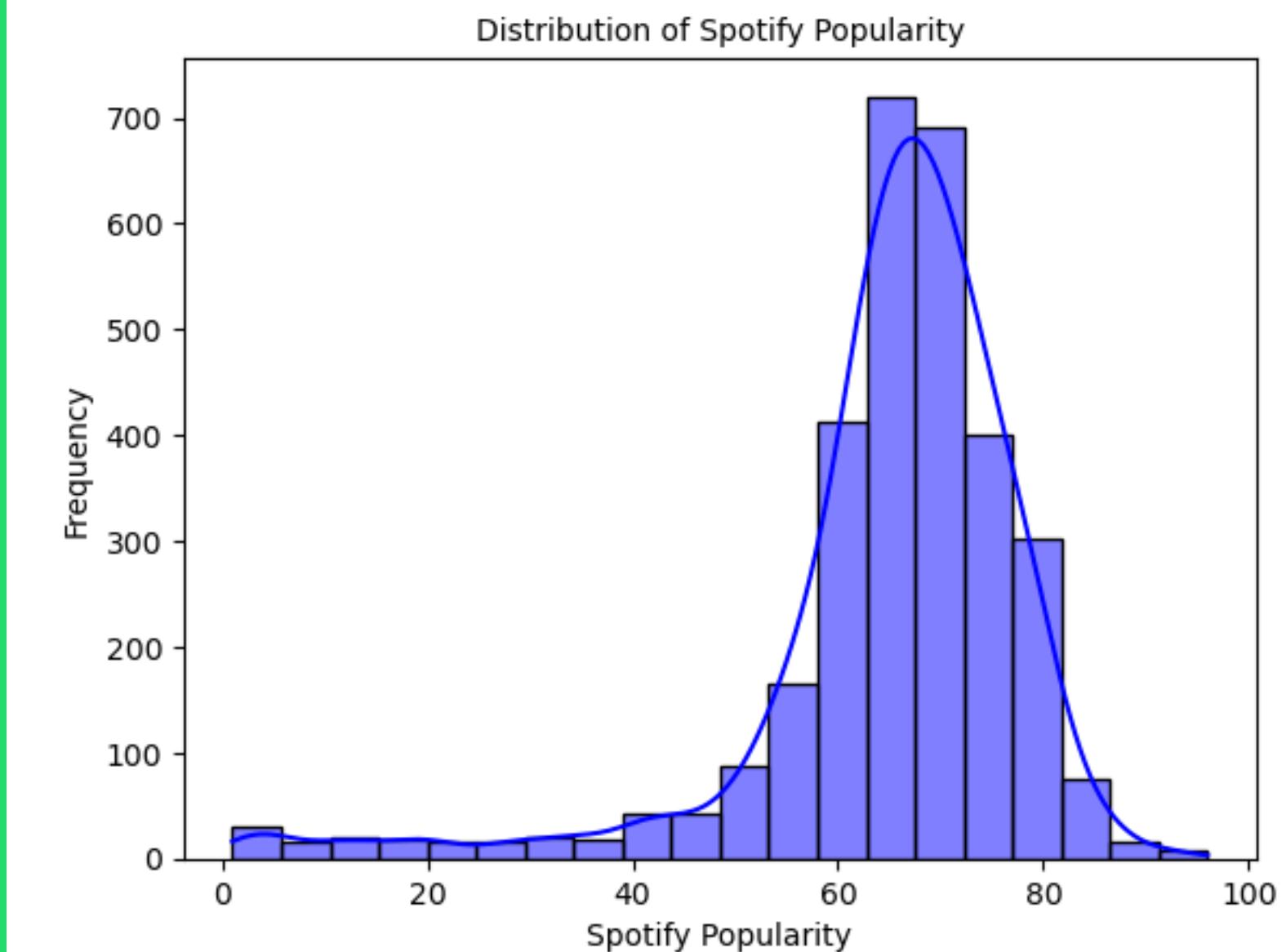
```
Q1 = data_cleaned['Spotify Popularity'].quantile(0.25) # First quartile (25th percentile)
Q3 = data_cleaned['Spotify Popularity'].quantile(0.75) # Third quartile (75th percentile)
IQR = Q3 - Q1 # Interquartile range
lower_bound = Q1 - 1.5 * IQR # Lower bound
upper_bound = Q3 + 1.5 * IQR # Upper bound

outliers = data_cleaned[(data_cleaned['Spotify Popularity'] < lower_bound) | (data_cleaned['Spotify Popularity'] > upper_bound)]

print(f'The lower bound is {lower_bound} and the upper bound is {upper_bound} in the dataset')
```

The lower bound is 45.5 and the upper bound is 89.5 in the dataset

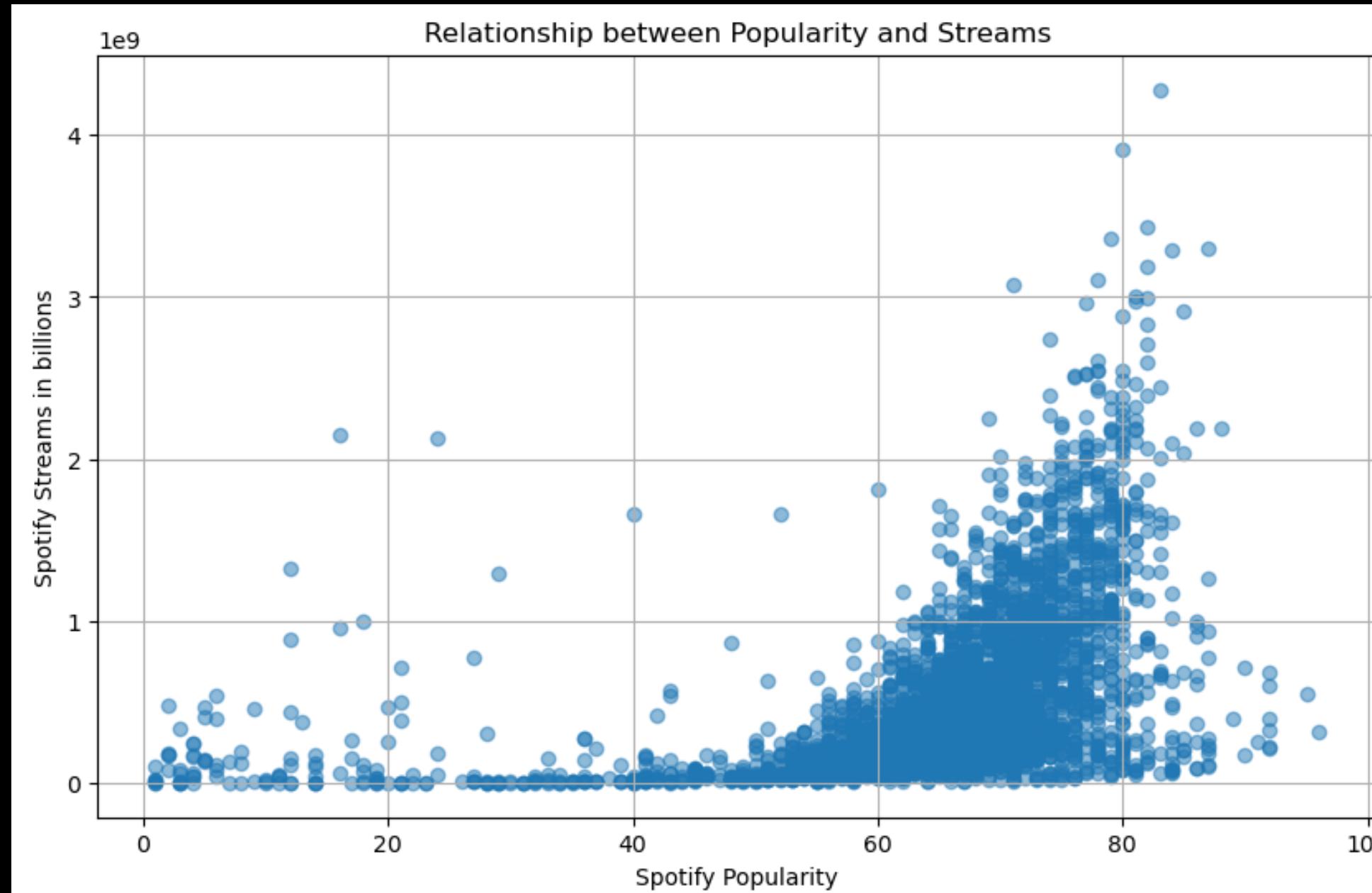
Tracks with popularity scores below 45.5 may indicate niche tracks or tracks with limited reach, while scores above 89.5 suggest that they are viral content.



Central Tendency: The distribution peaks around a Spotify Popularity score of 60–70, indicating that most tracks fall within this popularity range.

Shape of Distribution: The graph shows a roughly normal distribution, slightly skewed to the left. This suggests that while most tracks are moderately popular, there are fewer highly unpopular tracks.

Popularity vs Streams



Correlation between Popularity and Streams: 0.4444453162031849

Having a correlation of 0.44 means that it is a moderate correlation, meaning that the more streamed a song, the more popular it is. However, there are other factors that can influence popularity rather than only the number of streams (like social media presence or marketing).

Top Performers 2024

Artist	Spotify Streams
Taylor Swift	1.461244e+09
Benson Boone	1.175779e+09
FloyyMenor	6.750792e+08
Billie Eilish	6.704602e+08
Sabrina Carpenter	6.071973e+08
Artemas	6.013093e+08
Tommy Richman	3.904709e+08
Dua Lipa	3.596556e+08
Kendrick Lamar	3.237039e+08
Shaboozey	3.119285e+08

```
data_2024 = data_cleaned[data_cleaned['Release Date'].dt.year == 2024]

artist_streams_2024 = data_2024.groupby('Artist')['Spotify Streams'].sum().reset_index()

top_2024_artists = artist_streams_2024.sort_values(by='Spotify Streams', ascending=False).head(10)

top_2024_artists
```

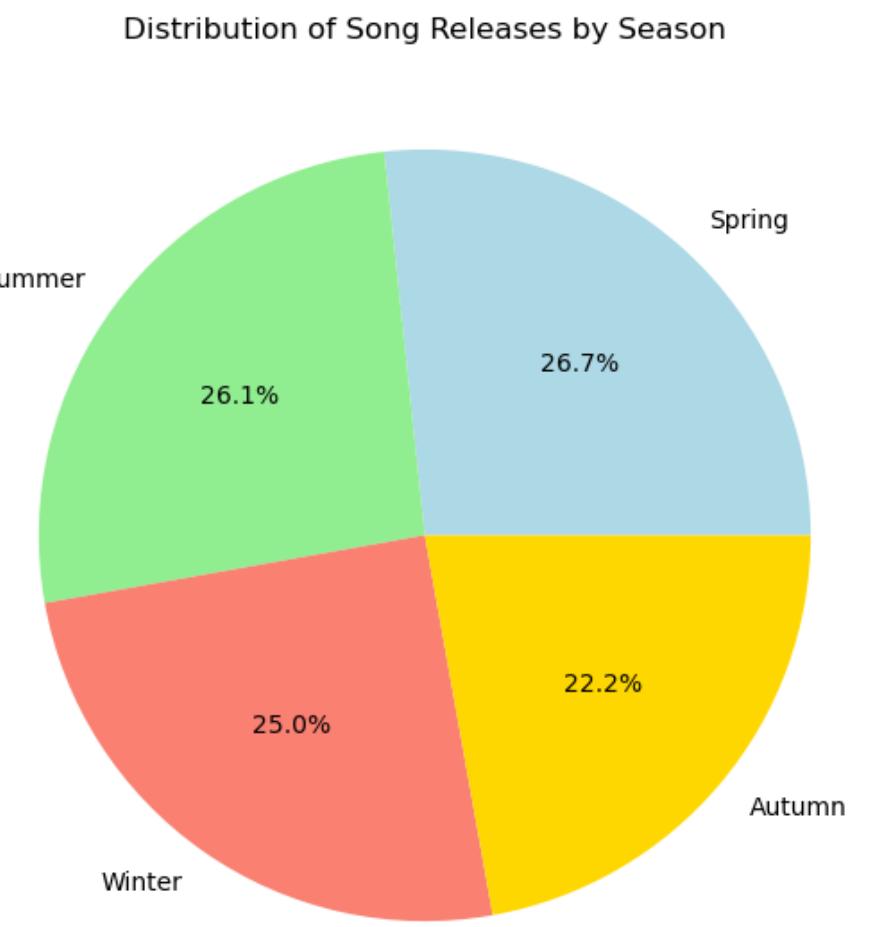
The most streamed artists that released at least a single in 2024. However, the Spotify Streams is an accumulative variable, meaning that the release date is relevant.

Seasonal Releases

```
data_cleaned['Year'] = data_cleaned['Release Date'].dt.year
data_cleaned['Month'] = data_cleaned['Release Date'].dt.month

def get_season(month):
    if month in [12, 1, 2]:
        return 'Winter'
    elif month in [3, 4, 5]:
        return 'Spring'
    elif month in [6, 7, 8]:
        return 'Summer'
    else:
        return 'Autumn'

data_cleaned['Season'] = data_cleaned['Month'].apply(get_season)
```



Top Performers 2024

```
artist_top_song_2024 = data_2024.loc[data_2024.groupby('Artist')['Spotify Streams'].idxmax(),  
                                     ['Artist', 'Track', 'Release Date', 'Spotify Popularity', 'Spotify Streams']]  
top_2024_artists_pop = artist_top_song_2024.sort_values(by='Spotify Streams', ascending=False).head(10)  
top_2024_artists_pop
```

	Artist	Track	Release Date	Spotify Popularity	Spotify Streams
5	Benson Boone	Beautiful Things	2024-01-18	86.0	900158751.0
6	FlooyMenor	Gata Only	2024-02-02	92.0	675079153.0
2	Artemas	i like the way you kiss me	2024-03-19	92.0	601309283.0
17	Sabrina Carpenter	Espresso	2024-04-12	95.0	547882871.0
12	Taylor Swift	Fortnight (feat. Post Malone)	2024-04-18	89.0	395433400.0
0	Tommy Richman	MILLION DOLLAR BABY	2024-04-26	92.0	390470936.0
1	Kendrick Lamar	Not Like Us	2024-05-04	92.0	323703884.0
24	Shaboozey	A Bar Song (Tipsy)	2024-04-12	96.0	311928522.0
335	Michael Marcagi	Scared To Start	2024-01-12	84.0	284119669.0
258	Natanael Cano	Madonna	2024-03-08	85.0	272838329.0

The Spotify popularity score is a metric that quantifies the popularity of a song or artist on Spotify.

This score is calculated based on various factors, including the number of streams, user interactions like saves and shares, and the track's inclusion in popular playlists.

A song could have high total streams but a lower popularity score if:

- The streams peaked in the past and have slowed down recently (according to when the data was retrieved).
- Listener engagement is lower (song or artist frequently skipped).
- It's no longer being actively recommended or added to playlists.

Most Streamed Artist 2024

Taylor Swift

1.5 Billion Streams

•••

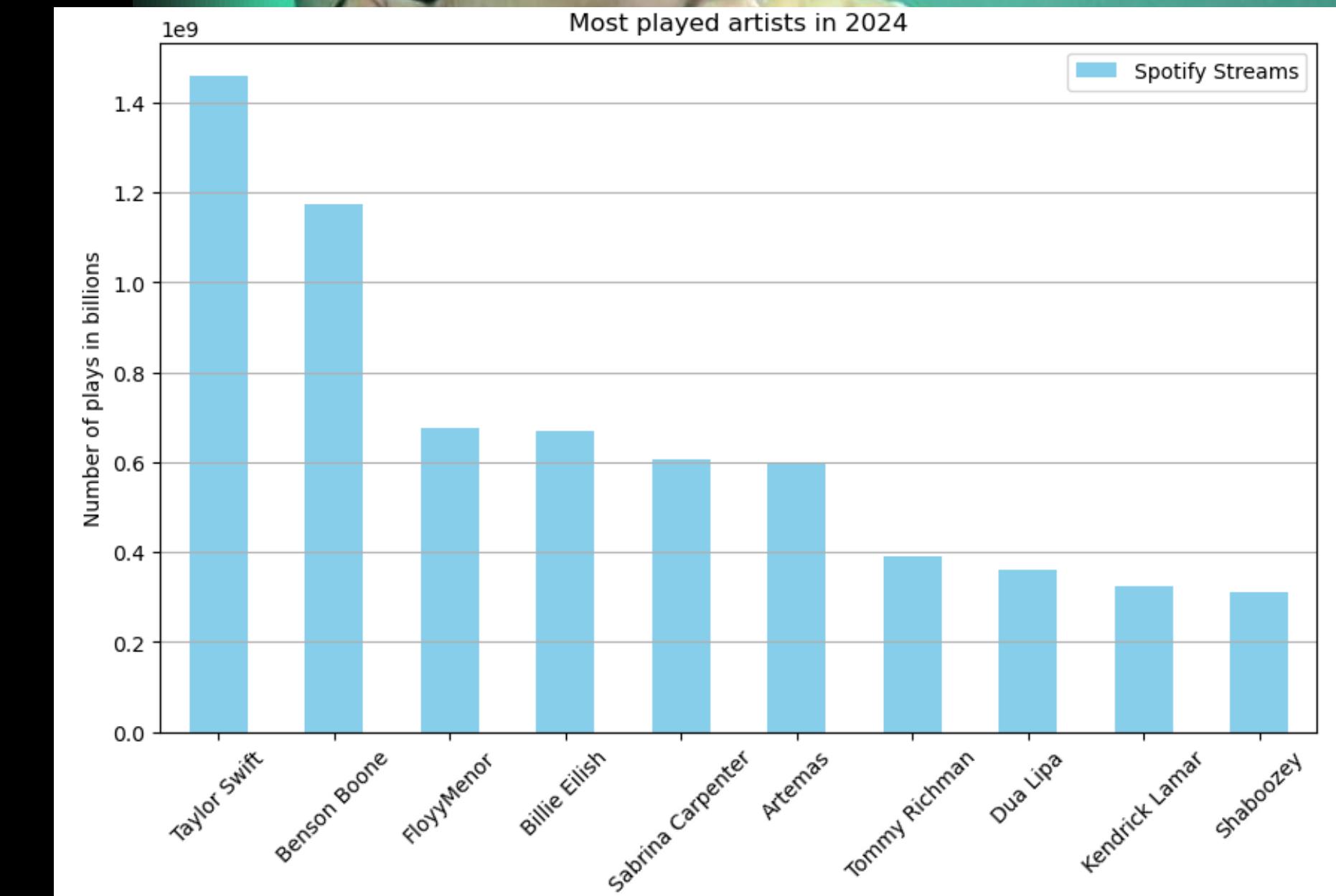
Popular Releases



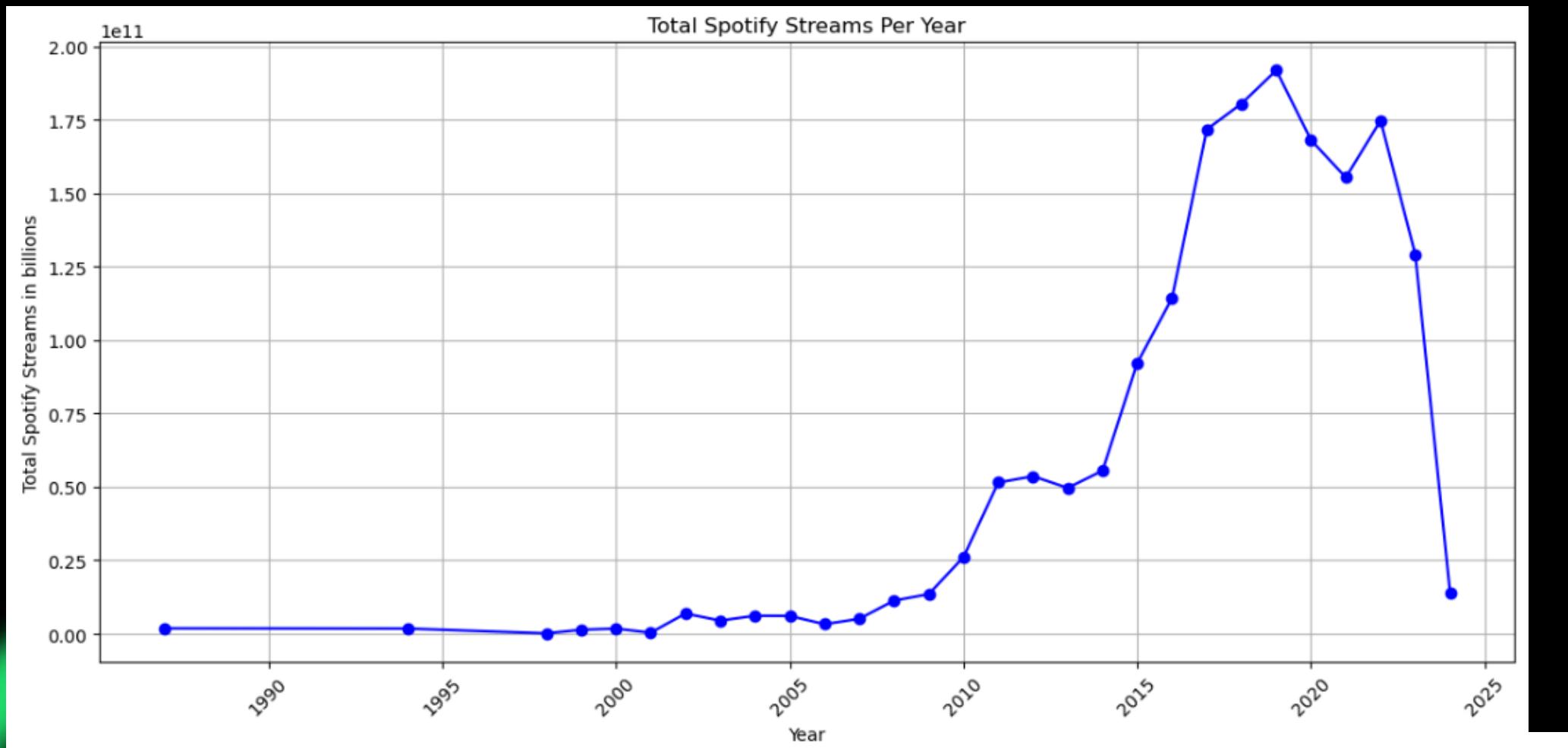
Her songs were not in the top 5 of most streamed songs or were not even rated as the most popular ones.



However, she had a massive Tour during 2024 which might have influenced the streams of the songs.



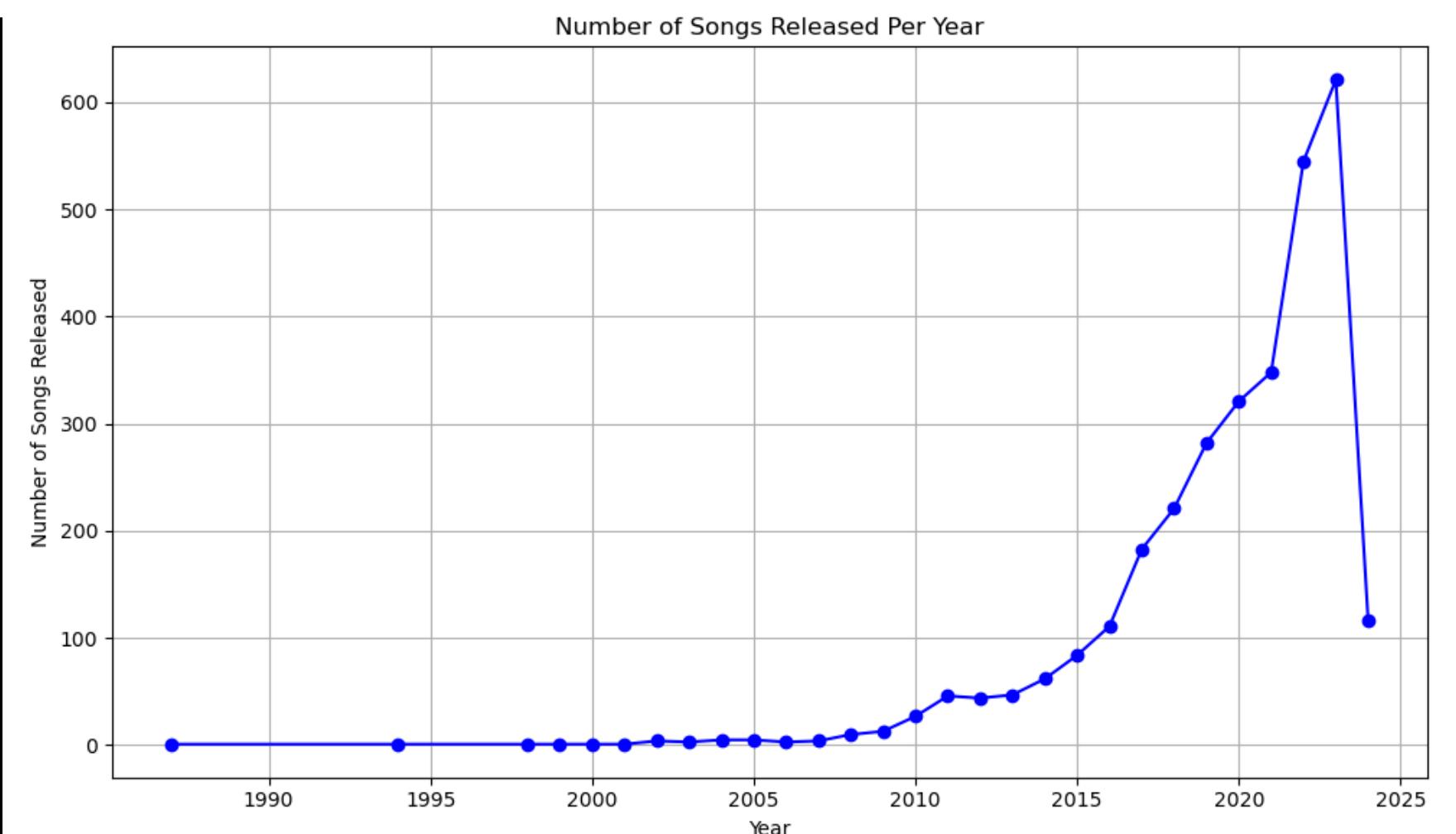
Trend Over Time



The Trend seems to be increasing, however in 2024 we have a massive sink. The initial approach was to think that 2024 was not a year with many music releases. So we study the 2024 release date.

The latest data retrieved is in June, which means that for 2024, almost half of the data is missing.

This will have an important impact to the analysis. However, it will be explained later.



Top Performers Over Time

All

Music

Podcasts

Audiobooks



1
Blinding Lights
The Weeknd



2
Shape of You
Ed Sheeran

```
artist_streams_over_time = data_cleaned.groupby('Artist')['Spotify Streams'].sum().reset_index()
most_streamed_artists = artist_streams_over_time.sort_values(by='Spotify Streams', ascending=False)

most_streamed_artists.head(10)
```

	Artist	Spotify Streams
123	Bad Bunny	3.493851e+10
1198	The Weeknd	3.332461e+10
333	Drake	3.004276e+10
1172	Taylor Swift	2.810369e+10
353	Ed Sheeran	2.401443e+10
987	Post Malone	2.393399e+10
82	Ariana Grande	2.158340e+10
602	Justin Bieber	1.664780e+10
153	Billie Eilish	1.632367e+10
814	Maroon 5	1.600031e+10

```
song_and_artist = data_cleaned[['Track', 'Artist', 'Release Date']]
song_streams_over_time = data_cleaned.groupby('Track')['Spotify Streams'].sum().reset_index()
merged = pd.merge(song_streams_over_time, song_and_artist.drop_duplicates(), on='Track', how='left')

most_streamed_songs = merged.sort_values(by='Spotify Streams', ascending=False)
|
most_streamed_songs.head(10)
```

	Track	Spotify Streams	Artist	Release Date
351	Blinding Lights	4.281469e+09	The Weeknd	2019-11-29
2269	Shape of You	3.909459e+09	Ed Sheeran	2017-01-06
1053	Happier	3.466207e+09	Marshmello	2018-08-16
1054	Happier	3.466207e+09	Ed Sheeran	2017-03-03
2363	Someone You Loved	3.427499e+09	Lewis Capaldi	2018-11-08
2450	Sunflower - Spider-Man: Into the Spider-Verse	3.358704e+09	Post Malone	2018-10-18
173	As It Was	3.301815e+09	Harry Styles	2022-03-31
2390	Starboy	3.291262e+09	The Weeknd	2016-09-22
1853	One Dance	3.192204e+09	Drake	2016-04-04
1417	Let Me Love You	3.118201e+09	Mario	2004-01-01

Key Points

User engagement

Top 5 most ranked artists on Spotify

Most streamed songs by top 5 artists on Spotify

Top 5 artist and most streamed songs

YouTube

User engagement comparison on Spotify

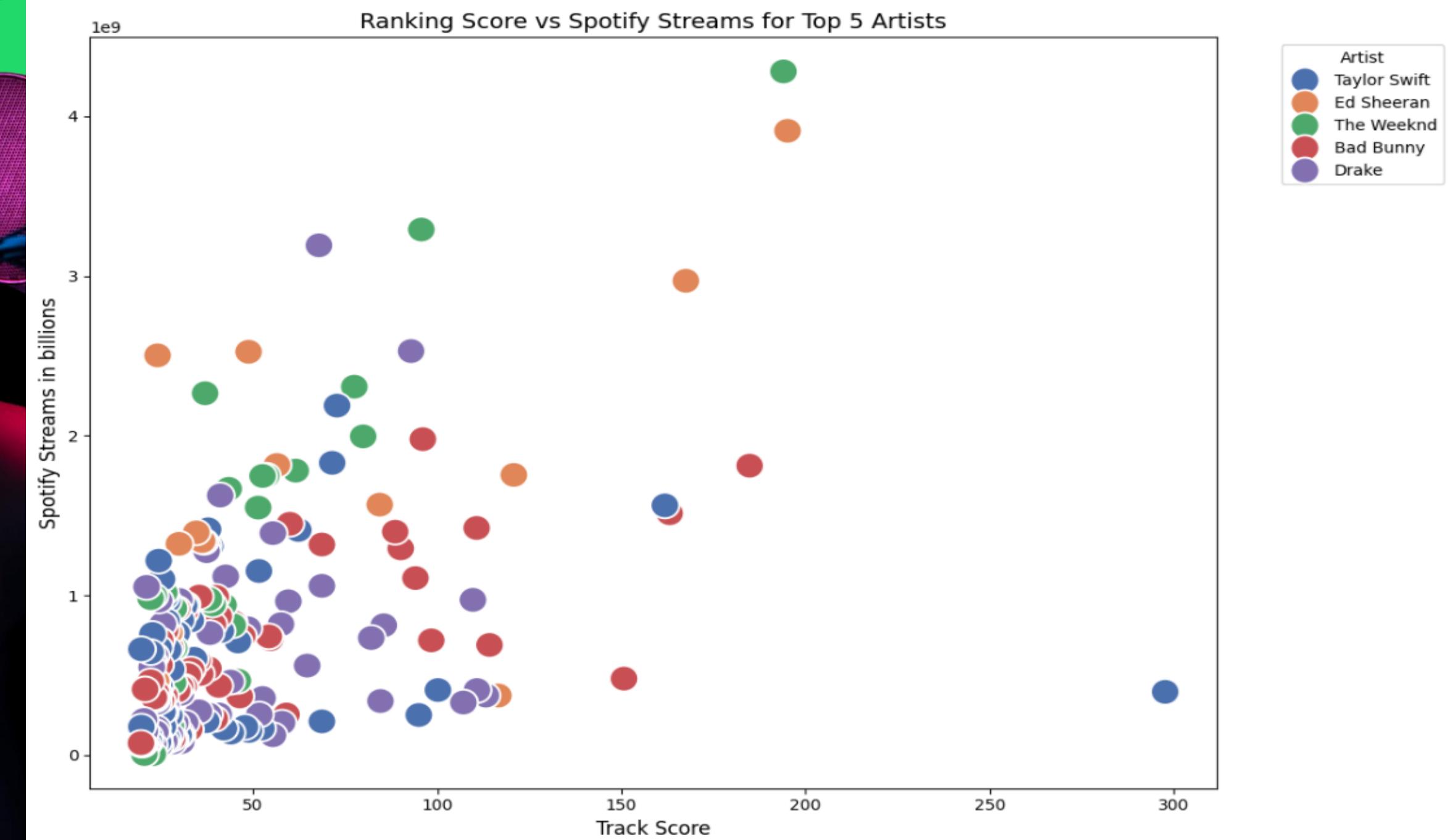
YouTube, TikTok

Seasonal Trends impact on ratings

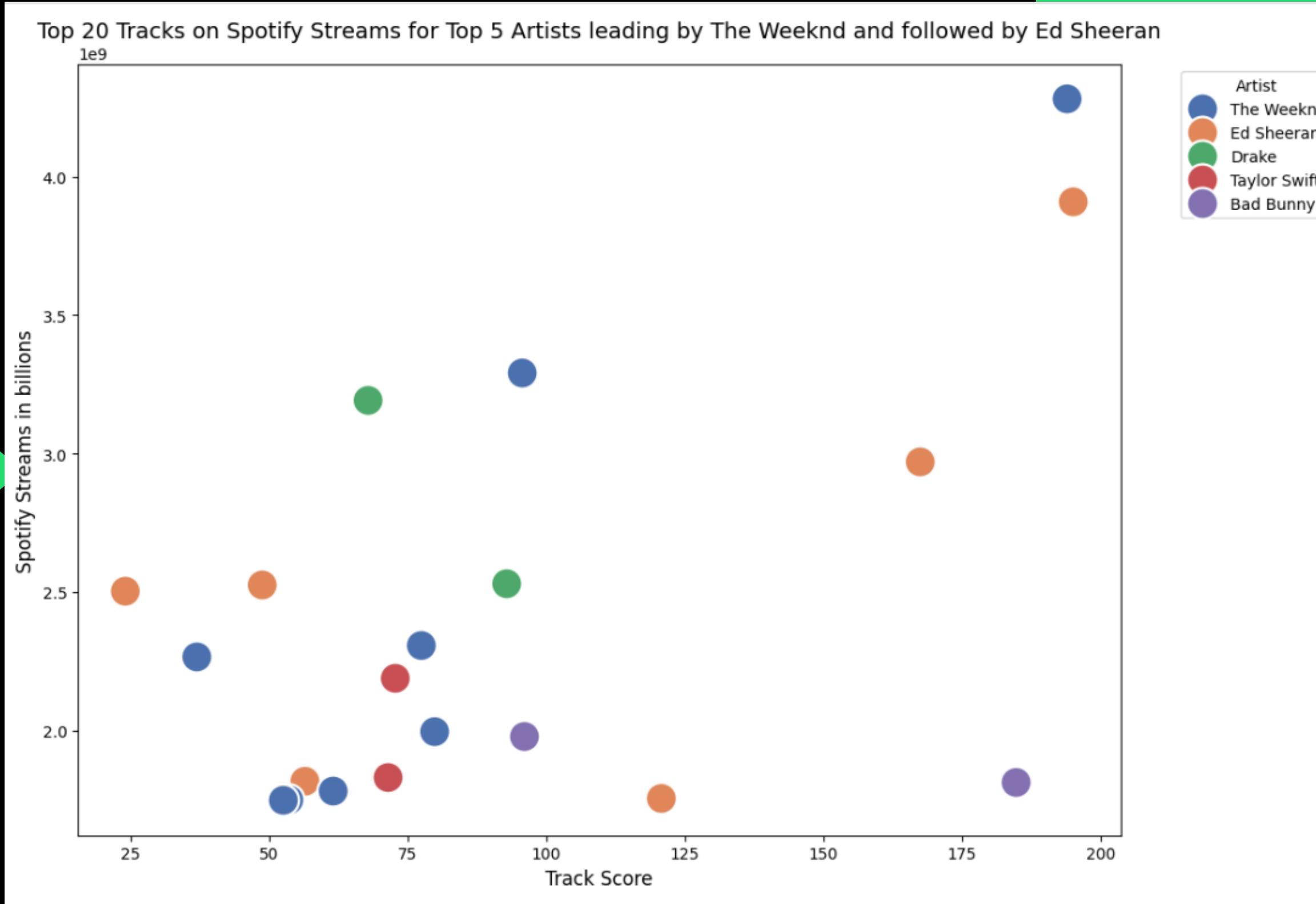


Top 5 Artists on Spotify

- Taylor Swift
- Ed Sheeran
- The Weeknd
- Bad Bunny
- Drake



Top 20 Tracks on Spotify



The Weeknd leads with 4B+ streams, followed by Ed Sheeran with ~4B streams

Drake with ~3B streams is strong competitor

No strong correlation - some high scoring tracks have lower streams and vice versa

Most Played Tracks on Spotify

The Weeknd - Blinding Lights 4B+ streams

Ed Sheeran - Shape of You ~4B streams and high ranking

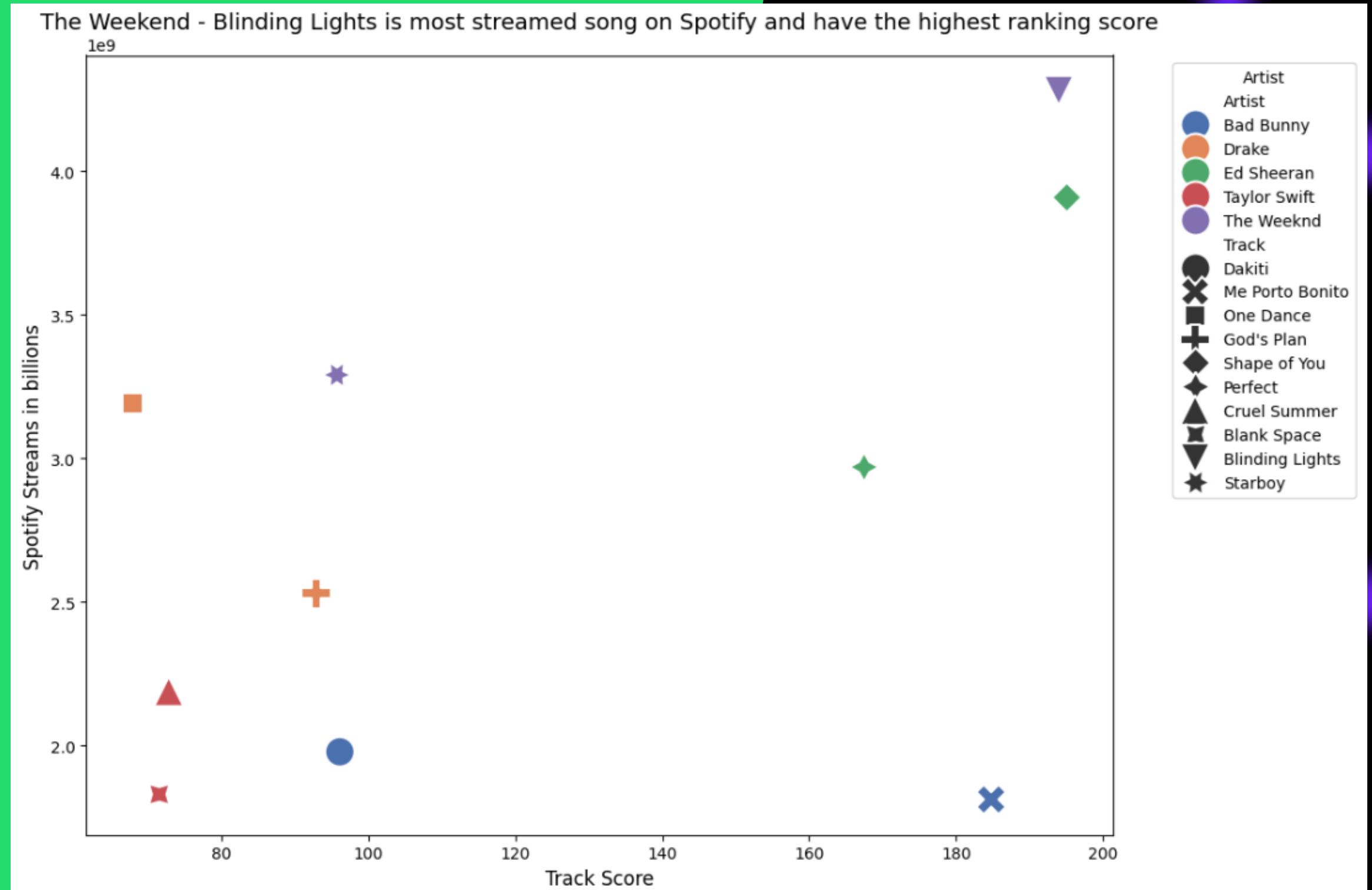
Drake - One Dance

The Weekend - Starboy

Ed Sheeran - Perfect 3B+ streams

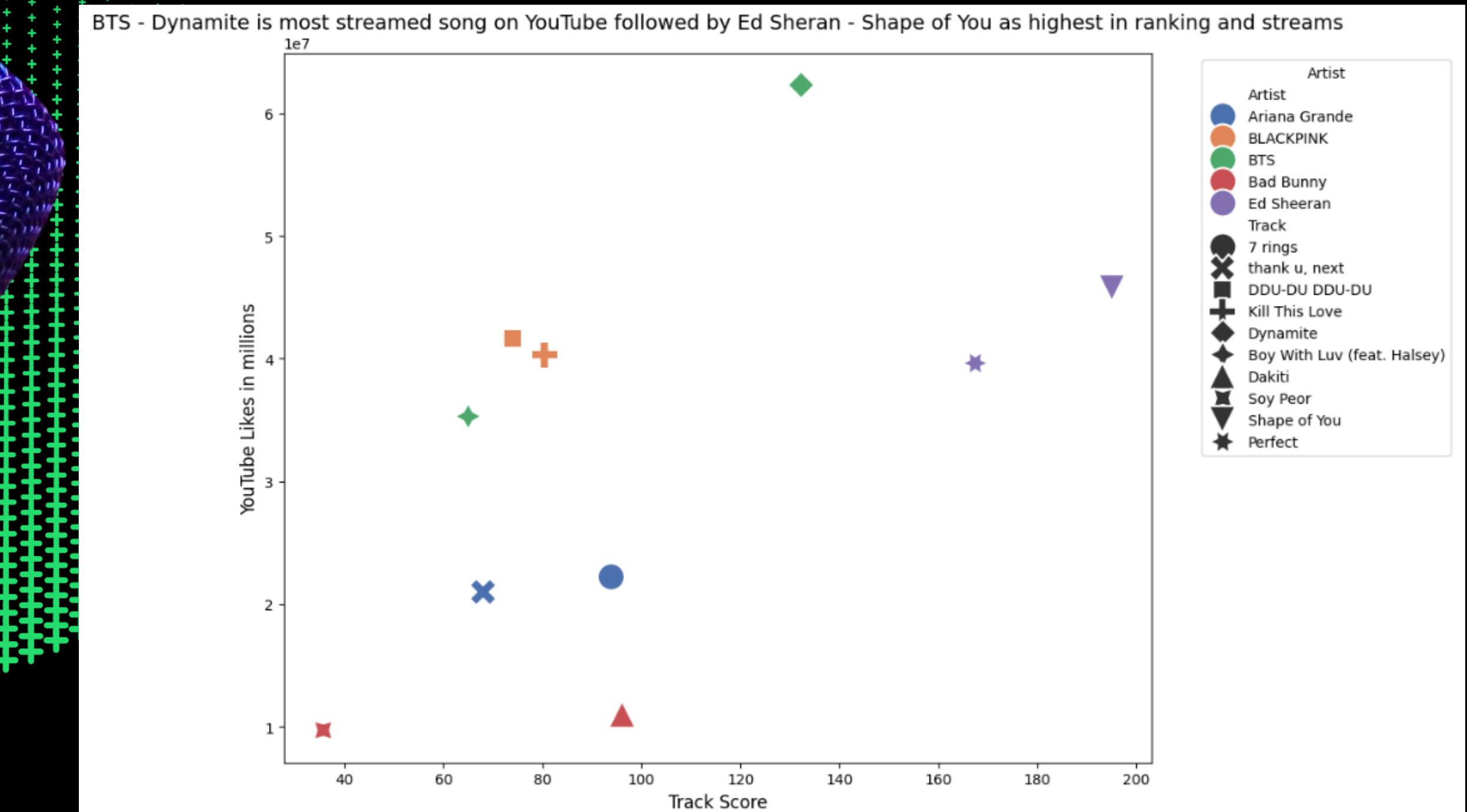
Bad Bunny - Me Porto Bonito high rank/lower score

Taylor Swift >2B streams, but lower ranking



Top 5 Artists on Youtube

- Ariana Grande
- Blackpink
- BTS
- Bad Bunny
- Ed Sheeran



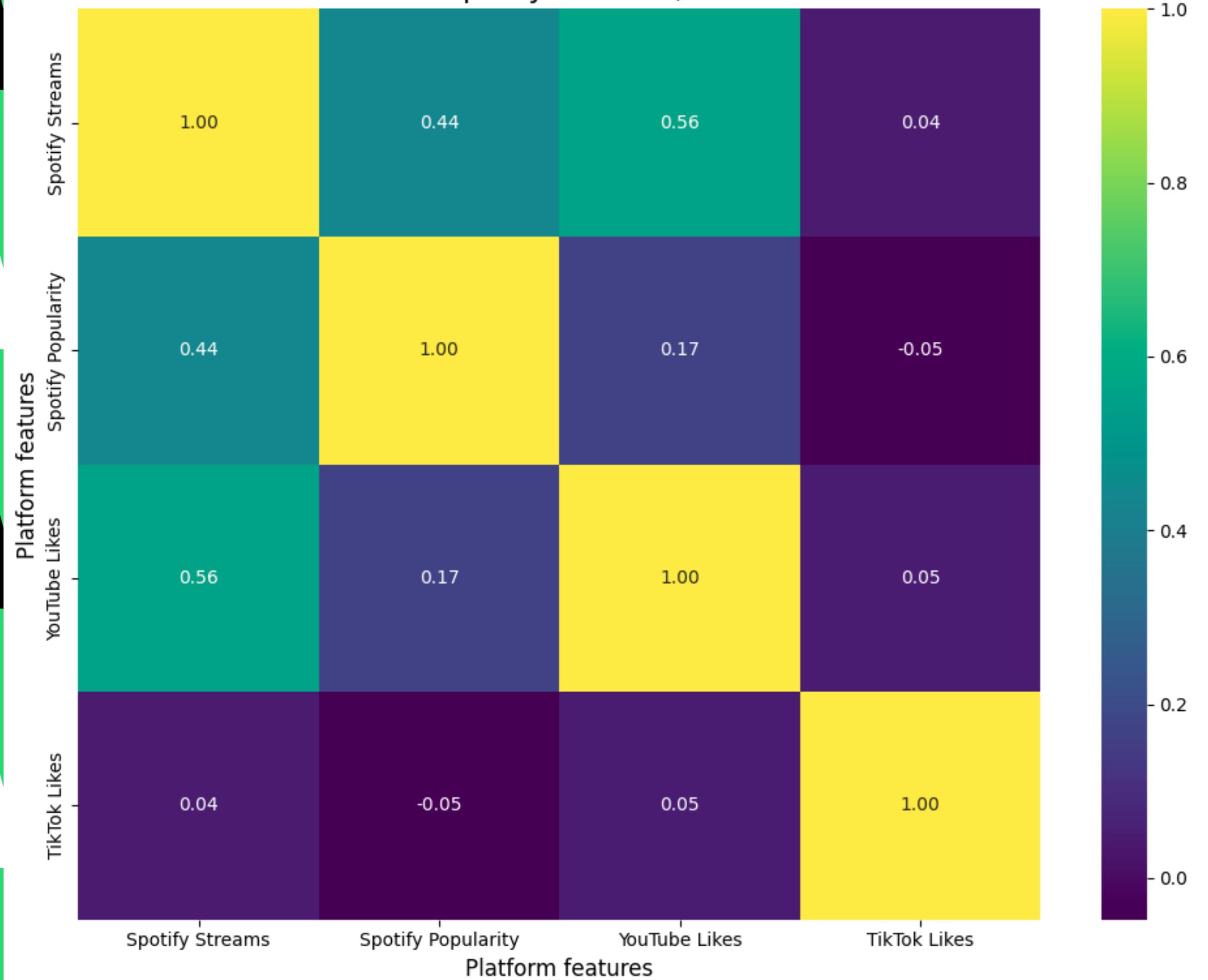
Most Played Tracks

Blinding Lights by The Weekend on Spotify with over 4 billion streams

Dynamite by BTS on YouTube with over 6 million Likes



Correlation Matrix of Spotify Streams, YouTube and TikTok



Spotify Streams correlates strongly with Spotify Popularity

YouTube and TikTok show strong correlation between views and likes

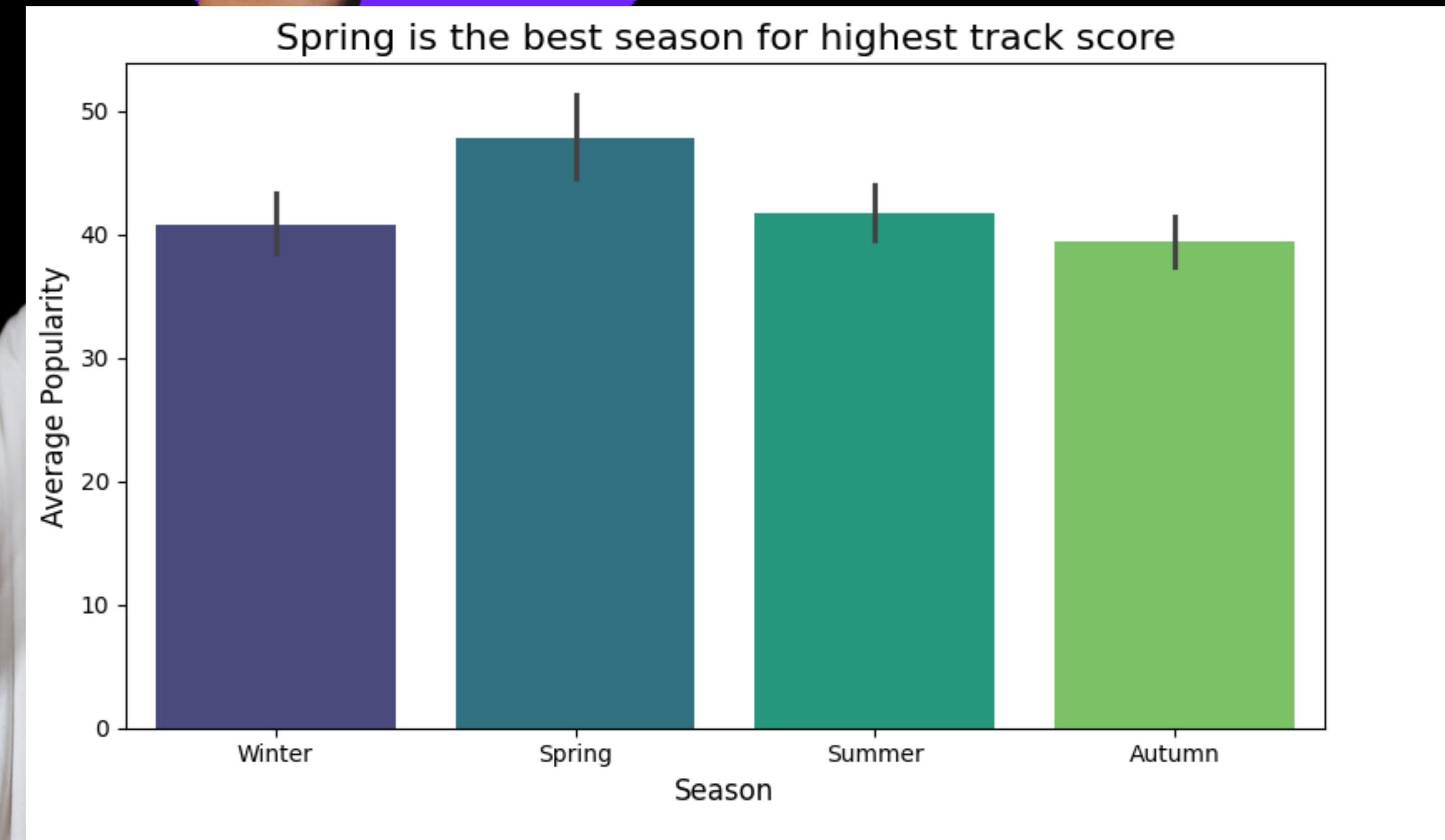
Positive correlation between Spotify and YouTube - when one increases does the other

Many users listen music on YouTube , contributing to this trend

Seasonal Trends



- Influencing factors:
- Psychological
 - Seasonal
 - Industry Driven



WRAPPED

Distribution of Song Popularity

Insight: Shows whether popularity is evenly distributed or if a few songs dominate.

Business Decision:

1. Diversify promotional strategies to support mid-tier artists.
2. Advertisers may target ads around popular songs, leveraging their visibility.
3. Adjust algorithms to balance new music discovery.

WID ADDER

Correlation Between Streams & Popularity

Insight: Identifies top-performing songs and artists.

Business Decision:

1. Labels should invest in multi-channel marketing (TikTok, YouTube, radio, playlists).
2. Emerging artists can still gain streams if they go viral—even if “popularity” metric is low.
3. Spotify might rethink how they rank & recommend songs.

Recommendations



WRAPPED

Most Streamed Songs & Artists

Insight: Identifies top-performing songs and artists.

Business Decision:

1. Emerging artists can learn from trending styles & strategies of top performers.

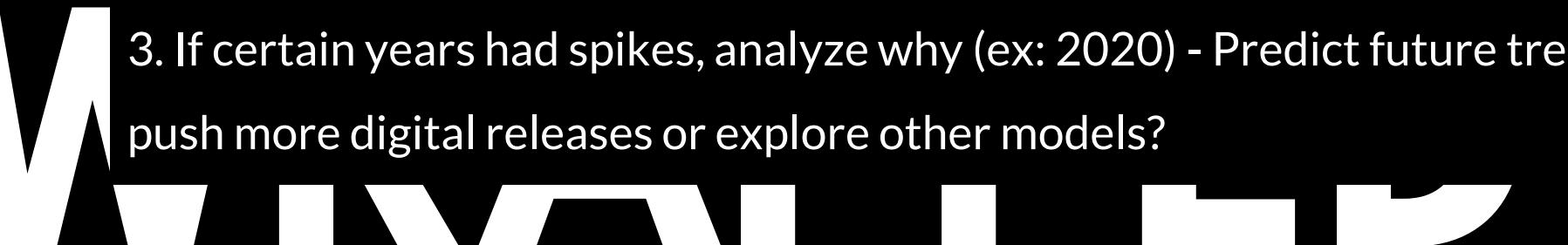
WRAPPED

Trend of Streams Over the Years

Insight: Tracks how streaming behavior has changed (e.g., peak years, growth rate).

Business Decision:

1. If streaming is increasing, platforms can monetize better (exclusive releases).
2. According to streaming Trends, search ways to acquire new clients/listeners (ads off platform).
3. If certain years had spikes, analyze why (ex: 2020) - Predict future trends: should labels push more digital releases or explore other models?



Recommendations



Comments

Comments

- The dataset does not have information of the whole year, affecting the study when 2024 is being analyzed.
- During the data cleaning process many rows were deleted, because of NaN values or special characters in Artist's Name or Track's Song, this reduced the dataset (around 30% of the dataset was deleted).
- The reduction might impact the results exposed during this EDA, but as time was a constraint instead of making some manual changes, we decided to delete this data.





Thank You

In case of doubt: Drop your questions!

In case of Feedback: Let us know your insight!