

Striking a Chord: Leveraging Spotify Insights to Forecast Hit Tracks for Musicians

Kiet T. Nguyen

Business Analytics, California State Polytechnic University, Pomona

GBA 6070 – Programming Foundation for Business Analytics

Professor: Fadi Batarseh

December 03, 2023

1. Introduction

In the world of making and listening to music, figuring out what makes a song really popular is a mix of art and science. Things like the beat, tune, speed, and other parts of the music are super important for a song to do well. My project aims to help musicians by looking at music and using data to understand what makes songs stand out.

2. Objective

By delving into the wealth of musical attributes within **Spotify's** dataset, my project aims to decode the winning elements that resonate with audiences. Through the synergy of data analysis and musical artistry, my findings will serve as a compass, empowering artists by analyzing successful *track features*, predicting *hit potential*, and offering *tailored guidance* to aid in the composition of commercially viable and popular music.

3. Dataset

Dataset: [Most Streamed Spotify Songs 2023](#) (last updated: August 26, 2023).

This dataset has lots of details about many different songs you can find on Spotify, Apple Music, Deezer, and Shazam. Each song in the dataset has a bunch of information about it, which helps us understand what the song is like and how well it does on these platforms. Important details in the dataset include:

- **Track Name:** Title of the song.
- **Artist(s) Name:** Name of the artist(s) responsible for the song.
- **Artist Count:** Number of artists contributing to the song.
- **Release Year/Month/Day:** Date of the song's release.
- **Spotify, Apple Music, Deezer, Shazam Playlists and Charts:** Metrics indicating the song's presence in Spotify playlists and its performance on Spotify charts.
- **Streams:** Total count of streams the song has amassed on Spotify.
- **Audio's features:** BPM (Beats per minute), key, mode, danceability, valence, energy, acousticness, instrumentalness, liveness, and speechiness.

4. Analytical questions

- What are the prevalent audio features that tend to correlate with a song's success or popularity?
- What's the correlation between specific artists' involvement or their attributes and the success of a song?
- How does the frequency of a song's appearance in music charts compare to its prevalence in curated playlists across various platforms?

5. Methodology

In this analysis, Python constructs like lists, dictionaries, for loops, and functions played crucial roles in handling and processing the data.

- **Lists:** Lists contained details like song names, artists, and various streaming metrics across different platforms. By using lists, it was easier to maintain and access these data points in an ordered manner, aiding in comparisons and calculations.
- **List comprehension:** list comprehension efficiently calculates and compiles the percentages of a track's presence on different platforms, facilitating their graphical representation for easy comparison.
- **Dictionaries:** A dictionary structure was utilized to track the number of hit songs per artist. This allowed for efficient counting and retrieval of artist-specific information, enhancing the analysis of each artist's impact on hit song counts.
- **For Loops:** For loops were extensively used to iterate through datasets and perform repetitive operations across multiple data entries. They were particularly crucial for iterating through song lists, streaming metrics, or artist data. These loops allowed consistent execution of operations on each song or artist, making it easier to analyze various attributes across the dataset.
- **Functions:** A function might have been created to calculate streaming statistics or to process artist collaborations. This modular approach streamlined the code, making it more organized and easier to maintain.

In short, these structures help us find answers to our main questions and make it easier to solve similar problems in other code.

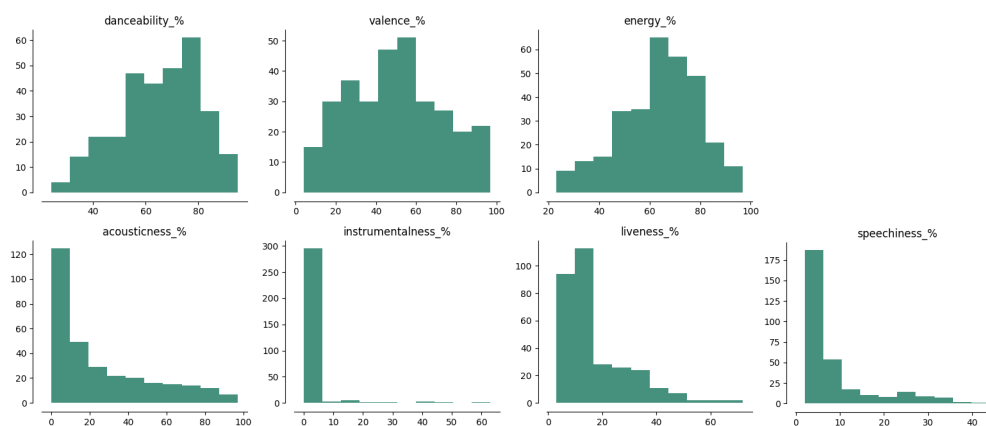
6. Exploratory Analysis

5.1. Music Analysis

At first, we're looking at different parts of music like how fast it is, how easy it is to dance to, how lively it feels, and other things to find patterns in songs that people like. But figuring out how many times a song needs to be played to be a hit can change a lot depending on different things. Generally, on big music apps like **Spotify** or **Apple Music**, a song usually starts getting noticed when it reaches hundreds of thousands or millions of plays. Assuming that songs with more than **500 million streams** are hits, we have defined that there're **309 hit songs** in the dataset. Now, we are going to do analysis on certain features that could affect the popularity of the song.

Audio Features Distribution

When we find out the most common things in songs that people like, we can help musicians make their future songs match what's popular now. This way, they can make sure their songs have the things that people really enjoy.

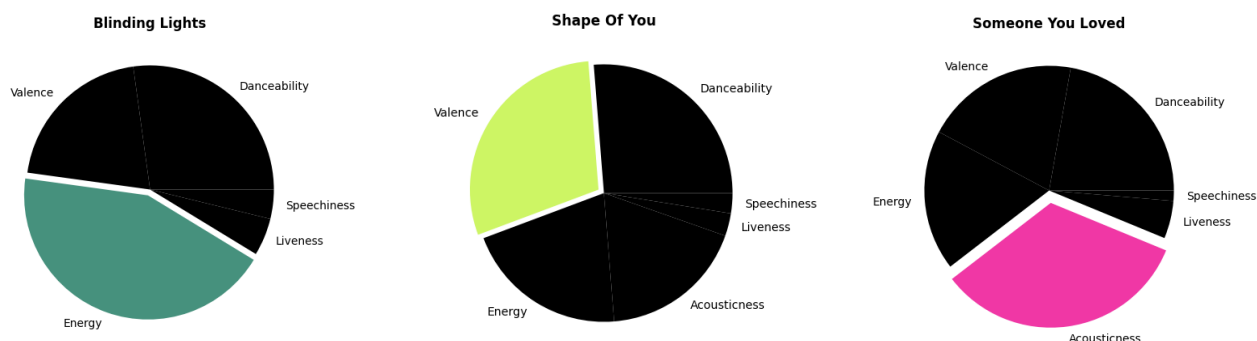


Audios' Features Distribution

When we look at popular songs, certain sound qualities matter more than others. *Danceability*, *Valence*, and *Energy* are really important in how songs are made. What's surprising is that even though things like *Acousticness* or *Instrumentality* are less important in most songs, they show up twice as much compared to the ones that have more of the important traits. Basically, hit songs are a mix of different sound qualities working together. Some are really noticeable, while others, even if they're not as common, still make a big difference in making the music more interesting and varied.

Distribution of Features in Top Songs

Now, we'll check if the songs people listen to the most match what we discovered earlier. If they do, we might think about changing how much we use the usual features compared to the less common ones, so our songs fit this trend better.

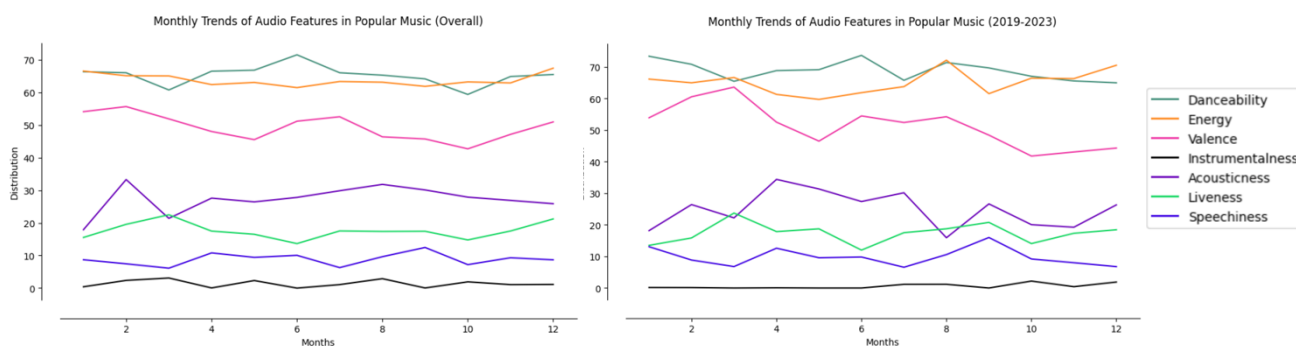


Audio's Features Distribution in popular songs

The study found that *danceability*, *positivity*, and *energy* are key in hit songs. In 'Someone You Loved,' despite its acoustic dominance, energy, positivity, and danceability are almost as important. However, there's little link between these elements and stream counts. Musicians can adjust these factors based on their goals or genres. These findings highlight specific features as potential key elements in varying proportions.

Evolution of Audio Feature Trends

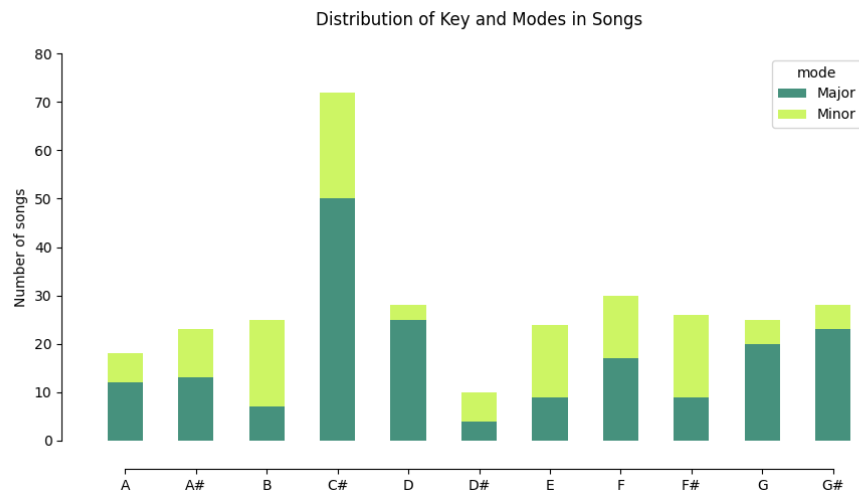
It's hard to know what people will like in music. But if we watch how certain things change in music over the years, we can guess what might become popular. This helps us change how much of those things we use in songs to match what most people might like.



Energy usually contrasts with *Danceability* but matches *Valence* overall. Sometimes, when *Energy* and *Danceability* match, *Acousticness* changes. Trends stay consistent, but small adjustments happen monthly to fit what's popular. Musicians can adapt their music to match audience preferences in each period.

Keys and Mode

When musicians use different keys and modes in their music, it can show what emotions they want to create. Major keys usually make songs feel happier, while minor keys can make them feel sadder or more thoughtful. Looking at which keys or modes are common in popular songs or certain types of music helps us see what emotions people like and why music is so successful or meaningful to them.



The stacked bar chart shows that **C#** is the most prevalent key, and within that key, the mode '**major**' holds a higher frequency compared to the mode 'minor'. This suggests a dominance of major tonality within the key of C# among the hit songs analyzed. Composing a song in C# major could potentially align with prevalent trends seen in successful music.

track_name	artist(s)_name	streams	key	mode
Blinding Lights	The Weeknd	3.703895e+09	C#	Major
STAY (with Justin Bieber)	Justin Bieber, The Kid Laroi	2.665344e+09	C#	Major
Watermelon Sugar	Harry Styles	2.322580e+09	C#	Major
Circles	Post Malone	2.132336e+09	C#	Major

Table 1: Songs that are composed in C# major mode.

It's evident that these songs have been major hits in recent years. Musicians might find it beneficial to incorporate these prevalent keys and modes into their new compositions.

Tempo

The speed of a song, called *tempo*, really affects how it feels. Faster tempos make it more lively, while slower ones make it calmer. Picking the right tempo is super important because it shapes how people feel about the music.

```
1 # Define the average tempo
2 average_tempo = hit_songs.groupby("released_year")["bpm"].mean().mean()
3 print(f"The average tempo of hit songs: {average_tempo} beats per minute.")
```

The average tempo of hit songs: 121.11601182680006 beats per minute.

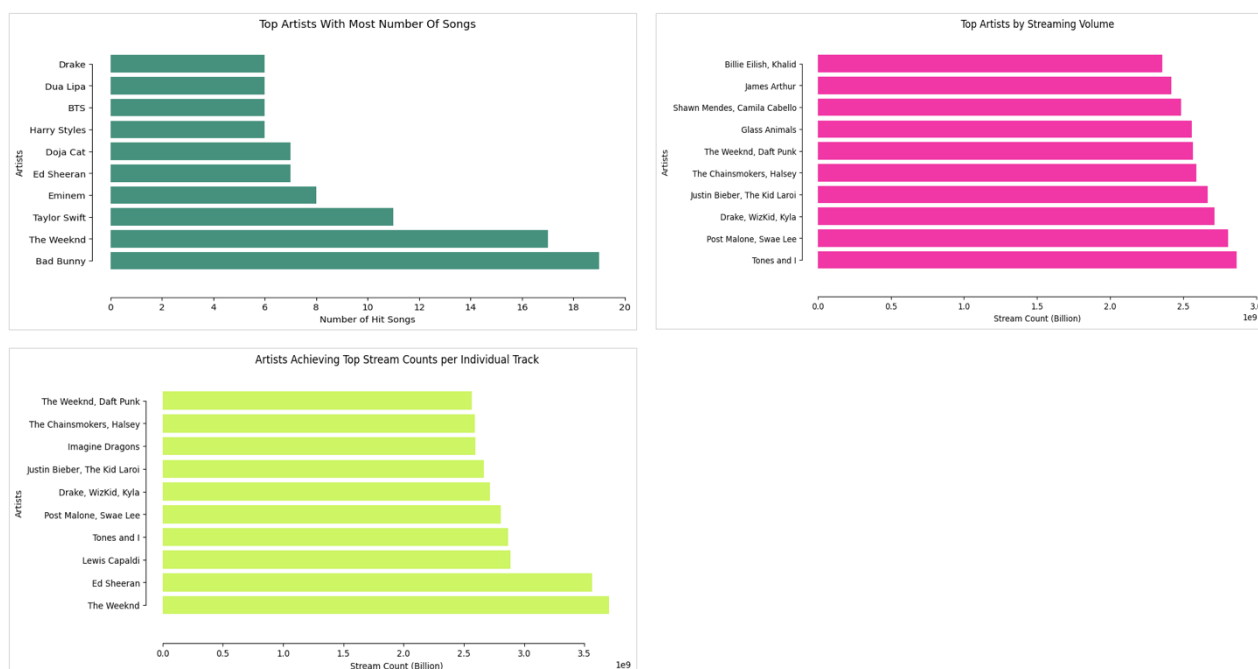
In these popular songs, the average tempo is **121 beats per minute (BPM)**, even the most streamed song, with over 3.5 billion listens, sticks around that pace. Some songs at 180 BPM got over 2.5 billion listens, but they're not as common.

Composers who want lots of people to like their music might aim for a tempo around 120 BPM. That's a speed many people are used to, but still catchy for a big audience. But they should also think about the emotions they want to express, the music style, and what they want their song to be like, as we've seen before.

5.2. Artist Impact

Looking at how artists affect a song's success means checking things like how many times songs are played, how many hit songs an artist has, and how often they work with others. Songs with more than one artist might do better on charts, playlists, and get more listens, showing they might have a big effect on success.

We'll see which artists show up a lot in three areas: having lots of hit songs, getting the most listens per song, and having the most listens overall.



It's clear that **The Weeknd** does really well in all three lists. His work with *Daft Punk* especially boosts his success. This shows that teaming up with others might help make hit songs.

```
1 # Determine if there's a correlation between number of artists in a song and its streams
2 artist_count_corr = hit_songs[['artist_count', 'streams']].corr()
3 artist_count_corr[:1]
```

	artist_count	streams
artist_count	1.0	-0.094165

```
1 # Proportion of collaboration
2 collab_hit_songs = hit_songs[hit_songs["artist_count"] > 1]
3 total_collab_hit_songs = collab_hit_songs["track_name"].count()
4 total_hit_songs = hit_songs["track_name"].count()
5
6 collab_proportion = total_collab_hit_songs / total_hit_songs * 100
7 print(f"The proportion of collaboration: {collab_proportion:.2f} %")
```

The proportion of collaboration: 32.04 %

The number **-0.094** shows that having more artists in a song doesn't really change how much it's listened to. But about **32%** of hit songs involve collaborations, showing they're pretty common in successful tracks. Even though it's not strongly linked to how many times a song is played, collaborations are still a big part of hit songs. So, teaming up with others can really help make a song successful.

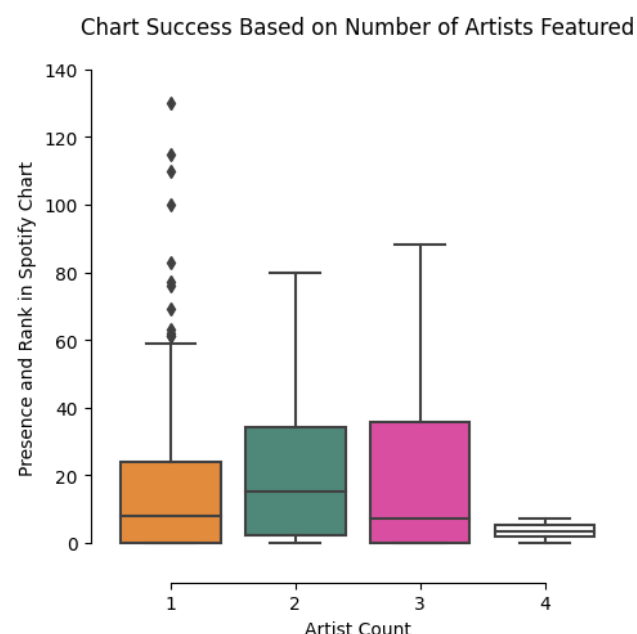
Top Artists with the highest number of collaborations

```
1 # Identify the top artists with the most collaborations
2 all_artists = pd.concat([collaborations_data["artist(s)_name"], collaborations_data['other_artists']])
3 all_artists.value_counts()[:10]
```

```
Bad Bunny      10
The Weeknd     10
Drake          6
Dr. Dre        4
Eminem         4
Dua Lipa       4
Post Malone    3
J Balvin       3
Doja Cat       3
Rauw Alejandro 3
dtype: int64
```

Several familiar names, including **Bad Bunny**, **The Weeknd**, **Eminem**, **Doja Cat**, **Dua Lipa**, and **Drake**, are present in both the top 10 artists with the most hit songs and those frequently engaging in collaborations. This intersection suggests a **potential relationship** between collaborating strategically and achieving success in terms of hit songs. Artists might purposefully collaborate to expand their visibility, reach wider audiences, or diversify their musical styles, thereby contributing significantly to their acquisition of hit songs.

Chart Performance



Also, songs by one artist have really high or really low positions on the charts. But songs with **two artists** working together usually do better overall, as shown by their middle position on the charts. However, songs with three artists, even though they usually land lower, have a wide range of positions on the charts. This shows that teaming up with others can make a big difference in how well a song does on the charts in the music world.

5.3. Cross-platform Performance

This investigation explores how songs perform across major streaming platforms like **Apple Music**, **Deezer**, **Spotify**, and **Shazam**. It examines the success of top songs on these platforms to understand audience engagement and trends specific to each service. This analysis compares *popularity*, *chart positions*, and *playlist impact*, unveiling how different audiences interact with music on each platform.

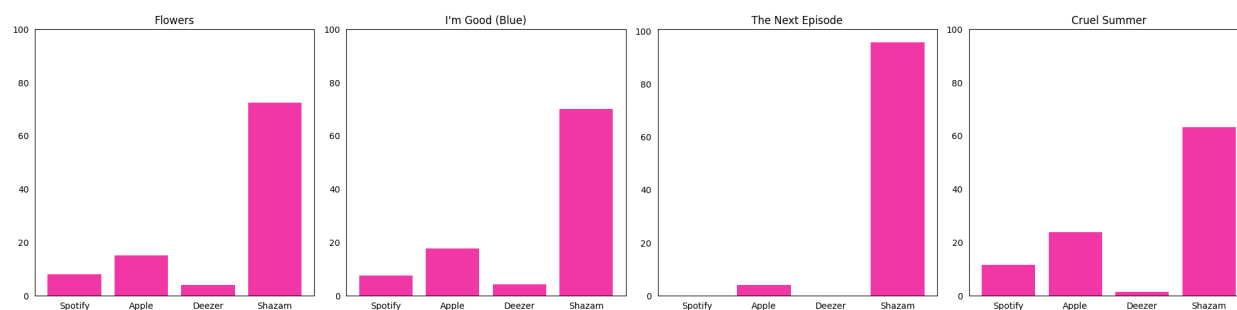
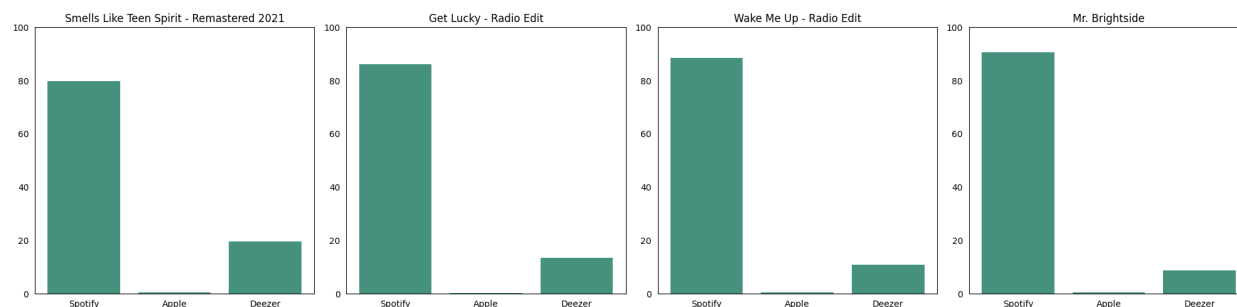


Chart Presence and Rank



Playlist Inclusion

Shazam is great for getting on the *charts*, while **Spotify** is good for being in *playlists*. If a song does well on both Spotify playlists and Shazam charts, it might mean it's both easy to find and keeps people interested for a long time. Musicians might want to make songs that aren't just catchy at first but also have something more that makes people want to listen to them again and again.

- **Catchy Right Away:** Songs that are easy to remember and have great parts catch people's attention right from the start. These things make people want to Shazam the song or put it in their playlists because they like it right away. Musicians can try to create strong, memorable parts or things that make their song different and catchy from the beginning.

- **Long-lasting Interest:** Besides being catchy, if a song stays on Spotify playlists, it might mean it has more to offer. This could be really good lyrics, complex music, emotions that people connect with, or sounds that you notice more the more you listen.

6. Conclusion

Analyzing the top songs on Spotify in 2023 tells us a few important things:

What Makes a Hit Song?

- Danceability, Energy, Valence, and Acousticness are crucial for a song to be a hit.
- Many songs use C# and have a "Major" vibe.
- Most songs have an average tempo of 121 beats per minute.

Artists Matter

- Songs with collaborations, especially those involving two artists, do better.
- Artists who work together a lot tend to do well on charts.

Different Platforms, Different Impact

- People quickly recognize songs on Shazam.
- Spotify encourages people to listen to songs repeatedly.

For future study:

- Check how different kinds of collaborations affect a song's success.
- Explore how each music feature separately affects a song's popularity.
- Understand how listeners behave on these platforms and how it influences a song's success.

Reflection

Conducting this project felt like the toughest one yet. I picked a topic after going through lots of references that inspired me. But then I hit a wall when I realized the dataset had too many issues, making my original idea unworkable. So, I went hunting for a new dataset, one I could handle with my current skills. I reworked my idea from there, which might not be the usual way, but I hope it helps me grow for future projects. I really tried to explore every part of the dataset, maybe even too much and I actually did not know if it was necessary. Fortunately, in the end, I got the result I was aiming for.

This project taught me a lot about *exploring data in detail*. I learned how important it is to understand the data well, dealing with missing information. Also, *starting with good data* is key, I learned the hard way that "garbage in, garbage out" holds true. As I experienced firsthand, errors can crop up later if the initial data isn't accurate. So, having well-prepared data streamlines the entire analysis process and prevents the need to backtrack to fix initial errors. It saves time and ensures accuracy in the analysis.

I also realized that *deciding if a song is a hit* isn't easy. It depends on how many times it's been streamed, its features, and how famous the artist is. But what makes a hit song can change from place to place, understanding music trends needs a good look at different things, not just one or two factors. It's like putting together a puzzle, considering lots of different pieces to see the whole picture.

References

Mackiek efer (2022). What makes a song become a HIT?

<https://medium.com/@maciek.fraczek85/what-makes-a-song-become-a-hit-c35593f08c9>

Phone Thit Htun (2020). ISSS608 Visual Analytics and Applications | Assignment 5

https://rstudiopubsstatic.s3.amazonaws.com/645408_6ecf0663fd5a40a9aec1bd51eb3246bb.html#2_dataset

Erik Herrström, Carlo Michelangelo Luetto, Rasmus Wangelin, William Oswin, Felipe Rocha, Chelsea Marie Alburger, Ellen Pai, April Pascu (2020). Spotify – A Decade Wrapped

<https://www.behance.net/gallery/93973343/Spotify-A-Decade-Wrapped>