

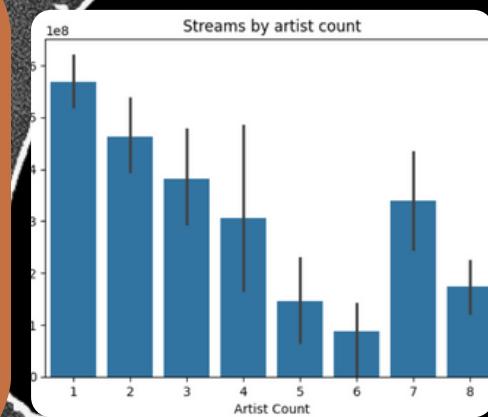
# PRODUCING YOUR NEXT HIT

## We feel you!

We know you dream of landing your next **viral** song every night. So, as data analysts, we are eager to uncover the 'secret recipe' for top hit songs on all platforms, so that we can help **YOU** make the next best **HIT**.

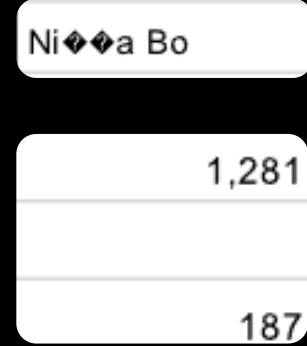
### Data Set

Our dataset contains **953 songs** with their artists, release dates, features, playlist and chart rankings on Spotify, Deezer, and Apple, BPM, key, mode, and various audio features like danceability, valence, energy, acousticness, instrumentalness, liveness, and speechiness.



### Data Problems

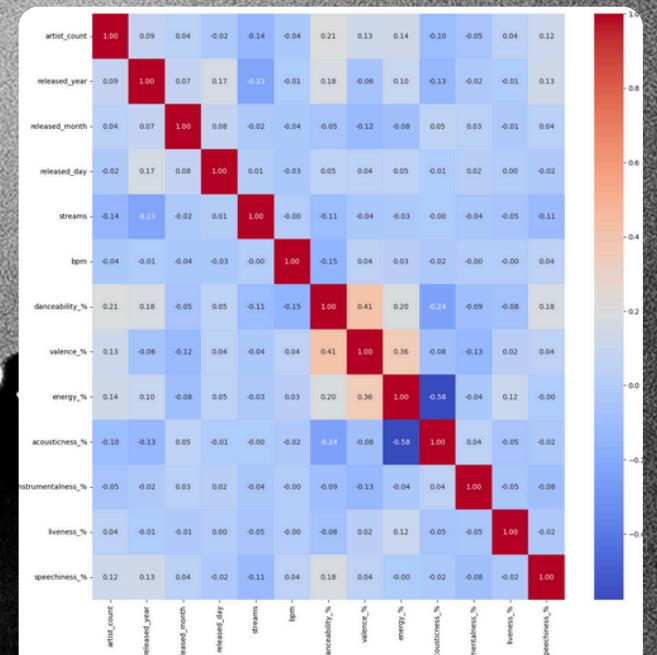
- Some song names were **not readable** by Pandas  
Soln: Python Script to replace all unknown characters
- There was **missing data** scattered around the data.  
Soln: Find the median of the data and fill in the empty cells
- There were **major outliers** that affected graphs.  
Soln: Use Python script to remove any outliers



### EDA

The released year has the **lowest correlation**, while the released day has the **highest correlation** with **popularity**.

Data distribution appears random, and charts and playlists are excluded because they **do not affect popularity**.



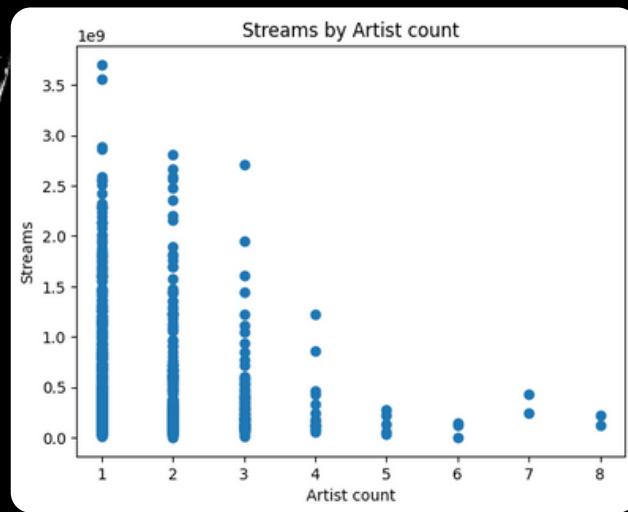
	Mean	Std	Median
artist_count	1.6	0.9	1
released_year	2018.2	11	2022
released_month	6	3.6	9
released_day	14	9.2	22
streams	514137400	566856900	673869000
bpm	123	28	206
danceability_%	67	14.6	96
valence_%	51	23.5	97
energy_%	64	16.6	97
acousticness_%	27	26	97
instrumentalness_%	1.6	8.4	91
liveness_%	18	13.7	97
speechiness_%	10	9.9	64

### So what makes a song popular?

Numerous factors can make a song **popular**. What are those **factors**? Fret not. We have done the hard work for you.

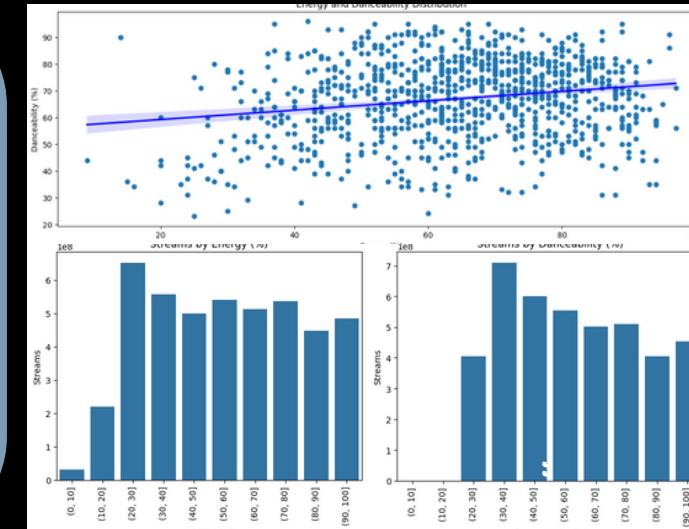
### Artist count

We suggest **limiting** collaborations to **two people** due to the **significant impact** it has on stream counts.



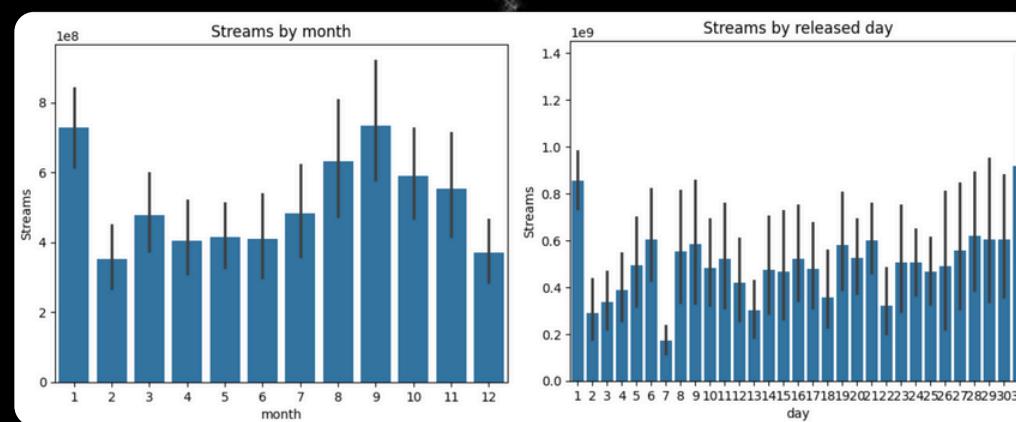
### Energy

The scatter graph reveals that danceability is **linked** to energy, thus enhancing your unreleased track's energy content can **significantly boost** streams.



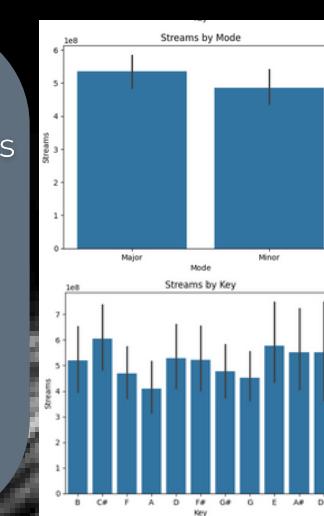
### Release Date

Songs released in **January** and **September** typically receive **higher streams**, suggesting that it's advisable to release your music on these dates.



### Keys

Upon close inspection, it is revealed a slightly higher trend in the **major key**, with **C# and E** being the top two trending keys, thus having your song in **C# or E major** the more euphonious choice.



### The magic numbers

What makes up the **best song**?

**1 Artist**  
Released in **Jan or Sept**  
Released on **31st or 1st**  
**BPM of 170-180**  
**C# Major or E Major**  
Dancability : **30% - 40%**  
Energy : **20% - 30%**  
Valence: **40-50%**  
Accousticness : **0% - 20%**  
OR **40% - 50%**  
Instrumentalness : **<70%**  
Liveliness : **60% - 70%**  
Speechiness : **4%**

### Multiple Linear Regression Model

How can we predict a song's **popularity** on the Spotify Chart?

### Linear Prediction Model

The model trained on all variables except playlists, charts, and streams to **predict** a song's Spotify Chart rank, but the prediction was inaccurate, resulting in **-1** on the charts.

### Polynomial Prediction Model

Instead of a linear prediction model, we decided to use a **curved line**, or a polynomial **prediction model**. The prediction was much more accurate, resulting in **62** on the charts.

**Linear**

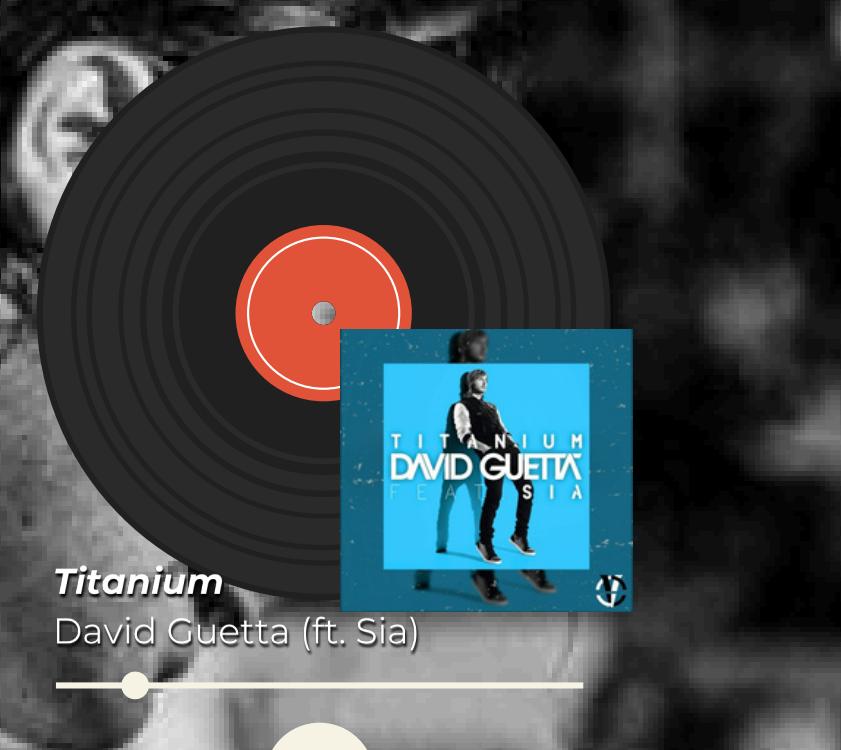
- R-squared: -0.018287426479190794
- Mean squared error: 336.04062713
- Root mean squared error: 18.33141094
- Intercept: -219.73044654680655
- Coefficients: [-0.29295538 ...]

Chart Rank  
-1

**Polynomial**

- R-squared: 0.32797893962858462
- Mean squared error: 345.53521478
- Root mean squared error: 18.58857754
- Intercept: -18159.883010747006
- Coefficients: [3.16247463e-09 ...]

Chart Rank  
62



The remix of "Titanium" was released on Sept 1 with high danceability, medium valence, high energy, and low acousticness, instrumentalness, liveliness, and speechiness. It has accumulated **1.56B** streams to date, making it a **perfect case study** for the **magic numbers**.