



Unlocking Global Rhythms

A Data Story on Spotify's Top Songs of 2024



Made for data enthusiasts everywhere

1,350 words • 11 pages



Overview

The Whys Behind the Project



Data

The Foundation of Our Story



EDA

Analysis: Unraveling the Music Mystery



Continental Trends

The Musical World Tour



Model

Predicting Song Greatness



Insights and Next Steps

Impact



Overview

The Whys Behind the Project



The Whys Behind the Project

Imagine you are a Data Analyst at Spotify, tasked with arming artists managers with the insights they need to conquer the charts. A new artist is on the rise, and to maximize their impact, we need to **understand the global music landscape, regional preferences, and the ingredients of a hit song**. This project is a deep dive into Spotify's 2024 data, uncovering the trends and patterns that drive success in the world of music.

To craft a winning strategy, I needed to answer critical questions:

- 1) What songs and artists are dominating the global stage?**
- 2) How do musical tastes differ across continents and where could our artist fit in?**
- 3) Can we predict a song's potential for success based on its sonic DNA?**

To be able to master this level of musical dissection, several tools were utilized. For all EDA related tasks, R was my pick. We all know and love the charts Spotify puts together every day to list our favorite music, and guess what? Charts can also be found here, only in a slightly different way. And to create them, Tableau was my visualization tool of choice. If you want to check out the graphs on Tableau Public, I'll leave the link [here](#): [Spotify Charts](#).

A bad song can make or break a playlist, the same way that choosing the wrong tool can severely impact the outcome of a model. That's why I chose to develop a prediction model using Python within the Jupyter Notebook environment. I utilized Python's powerful libraries, such as scikit-learn and Pandas, along with matplotlib.

Data: The Foundation of our story

I leveraged a rich dataset from Kaggle, a treasure trove of Spotify data spanning **72 countries** that includes the Global Top 50 as well.

This dataset, with over **1.4 million rows and 25 columns**, provided a snapshot of daily top 50 charts from **October 2023 to November 2024**.

This granularity allowed me to dissect trends not just globally, but also at a continent level - crucial for tailoring an artist's regional strategy.

Do you want to get a pick to the dataset information? I'll leave you a variable overview at the end of the document and here is the link to Kaggle: [Top Spotify Songs in 73 Countries \(Daily Updated\) | Kaggle](#).

Data (*Structuring and Cleaning's Version*)

Artists, producers, audio engineers, they all work together to make our favorite songs sound amazing. They achieve this by bringing that creative musical spark, but also by carefully erasing any trace of noise that could compromise the structure of the song. You could say they leave the audio squeaky clean. So here, my mission was to follow their steps and leave the dataset as clean as possible.

The dataset wasn't perfect; it had some missing values. The 'album_name' column had the most gaps. Since these missing values represented less than 0.05% of the data, I carefully removed them to ensure the integrity of my analysis.

There were NAs in the 'country' column, too. These values, however, were not removed, but changed. From the overview on the Kaggle dataset, the NAs on this column represent the Global Top 50, so I changed them to "GO" to keep using the same format the rest of the columns have.

The repeat song button is saved for very special occasions, and for very special songs. Duplicating the experience is something we all can appreciate, but for a dataset, duplicates are not as well accepted. Luckily for us, when I searched for them, I came up empty.

I spotted some outliers in the data, particularly in 'popularity' and 'duration_ms.' While it was tempting to discard them, I decided to keep them. In the music world, viral sensations and unusually long tracks can significantly impact trends. Removing them would have meant missing out on valuable insights into what makes the music landscape so dynamic.

Analysis: Unraveling the Music Mystery

Before diving into the big questions, I needed to get acquainted with the data.

I started with **descriptive statistics** to understand the **distribution** and possible **correlations** of key variables like popularity, danceability, and energy. This bird's-eye view revealed that:



Most songs in the dataset are quite popular with a mean of 76.4



Descriptive Statistics

Which makes sense, given they're from the top charts.



Something I found interesting is that both danceability and energy have a similar mean



Descriptive Statistics

It tells us their numbers are alike. Would it be that they are correlated? We'll find out.



The popularity histogram tells an interesting story. It's left-skewed, meaning most songs have high popularity scores



Distributions

This confirms that I'm looking at a selection of successful tracks, which is great because it will allow me to analyze what contributes to that success.



The histograms further proved how similar danceability and energy are



Distributions

I'll put on my detective hat to further investigate if they are correlated. Stay tuned!



The scatterplot of danceability/energy revealed a positive connection – songs that make you want to move often pack a high-energy punch



Correlations

This is a key insight for our artist: creating music that's both danceable and energetic could be a winning formula.



From danceability and energy compared to popularity, the focus should be on danceability



Correlations

It shows a more condensed correlation with popularity.

The World Tour

Now, let's embark on the musical world tour! I segmented the data to uncover regional trends, comparing top songs, artists, and musical features across continents. This is where things get really interesting – it's not a one-size-fits-all world when it comes to music.

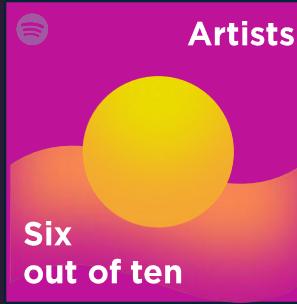
Europe Mix



Songs from Europe represent the 39.7% of all the data.



From the Top 10 Songs, 7 were on the Europe Top 10 as well.



From the Top 10 Artists, 6 were on the Europe Top 10, too.



Songs you lov

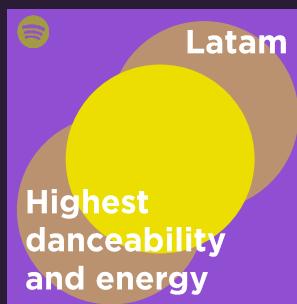
Continent Mix



Out of the Top 10, 2 songs are in Spanish.



KAROL G, Jimin, and Feid are part of the Top Ten Artists.



Latin America likes happy and upbeat songs the best.



Your past favo



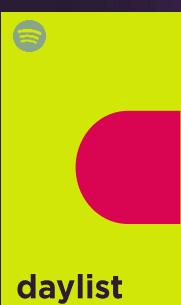
Oceania has six out of the Top Ten Artists on their charts, too.



Asia doesn't prefer explicit language.



Both continents like songs with a lot of lyrics.



Your day in a

The Top of the Top



Top Song Beautiful Things

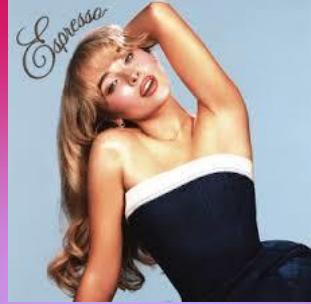
Number of repetitions

10,212

0.7% of the dataset



SPOTIFY.COM/WRAPPED



Top Artist Sabrina Carpenter

Number of repetitions

22,196

1.5% of the dataset



SPOTIFY.COM/WRAPPED

D

Your Library



Top 10 Song Features

Downloaded

↑↓ Recents



80

Danceability (Only two songs fall outside the middle 50%)

80% of the Top Ten songs fall in the most common range of values for danceability

70

Energy (Seven songs are not on the middle 50% of energy values)

70% of the Top Ten Songs are NOT in between quartile 1 and 3 of energy

70

Instrumentalness (Seven out of ten songs are on the middle 50% of values)

70% of the Top Ten Songs fall in the most common range of values

70

Mode (Seven songs have a mode of 1)

70% of the Top Ten Songs have the same mode

60

Speechiness (Four songs are on the middle 50% of values)

60% of the Top Ten Songs are NOT in between quartile 1 and 3

Predicting sound greatness

One of the key takeaways from The Top 10 Song features is that these songs don't show a clear differentiation when it comes to the analyzed audio characteristics. In other words, the information is still a little vague. So, to find more useful information, I took the liberty to work on a prediction model that will tell us what musical characteristics might make a song popular. Think of it as a 'hit song predictor'.

By feeding the model various audio features, I could estimate the likelihood of a song becoming popular. This model isn't a crystal ball, but it provides valuable guidance on what musical elements might resonate with listeners.

The selected features were: danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, time signature, and duration.



Key Audio Features

	Increased odds
1 Time Signature	464%
2 Valence	63%
3 Mode	48%
4 Loudness	38%
5 Duration	13%

From all the variables used on the model, **46%** (six out of thirteen) showed that for a one-unit increase on any of these variables, the odds of the song becoming popular would be as presented above



[SPOTIFY.COM/WRAPPED](https://www.spotify.com/wrapped)

Impact

I presented the numbers, but you must be thinking: okay, now what? And that question is indeed what's most important from all of this. What steps, insights and recommendations can we give the artist? Let's break it down.

- Tapping into the European market would be a smart move. I recommend starting there, given the sheer presence of European countries on the dataset
- Let's consider the importance of Latin American audiences, too. Breaking the language barrier when talking about English-speaking artists might be a little daunting, so my recommendation would be to secure collaborations with Spanish-speaking artists and marketing strategies targeted to Latin countries
- If the artist is interested in pursuing after the Asian market, there should a consideration on explicit language (explicit songs don't do well there)
- Keep in mind danceability and energy when creating a song, upbeat and happy songs do well (the Top Ten songs have average and above average values, respectively)
- A higher time signature shows as the best bet for song popularity. Given the dataset, a time signature of 4 would be the sweet spot, it provides versatility to a wide variety of tempos and styles, and even human physiology backs it up as it loosely associates with walking or a heartbeat
- Valence is another key feature to keep in mind, think of this one as the "happiness" of a song, or the musical positiveness conveyed by the track
- Having a mode of 1 is the way to go, this would mean a brighter and happier song

In summary: you, as the Data Analyst giving the ultimate advice to the artist's manager, should be aware of the European and Latin American market, and recommend more happy and upbeat songs to make the odds work in your favor. Don't forget authenticity and a relatable message to connect with the audience.

That's all for now. Thank you for tuning in!



Variable Overview

chr	spotify_id	The unique identifier for the song in the Spotify database	⋮
chr	name	The name of the song	⋮
chr	artists	The name(s) of the artist(s) associated with the song	⋮
dbl	daily_rank	The daily rank of the song in the top 50 list	⋮
dbl	daily_movement	The change in rankings compared to the previous day	⋮
dbl	weekly_movement	The change in rankings compared to the previous week	⋮
chr	country	The ISO code of the country of the Top 50 Playlist. If Null, then the playlist if 'Global Top 50'	⋮
Date	snapshot_date	The date on which the data was collected from the Spotify API	⋮
dbl	popularity	A measure of the song's current popularity on Spotify	⋮
Igl	is_explicit	Indicates whether the song contains explicit lyrics	⋮
dbl	duration_ms	The duration of the song in milliseconds	⋮
chr	album_name	The title of the album the song belongs to	⋮



Variable Overview

date

album_release_date

The release date of the album the song belongs to

⋮

dbl

danceability

A measure of how suitable the song is for dancing based on various musical elements

⋮

dbl

energy

A measure of the intensity and activity level of the song

⋮

dbl

key

The key of the song

⋮

dbl

loudness

The overall loudness of the song in decibels

⋮

dbl

mode

Indicates whether the song is in a major or minor key

⋮

dbl

speechiness

A measure of the presence of spoken words in the song

⋮

dbl

acousticness

A measure of the acoustic quality of the song

⋮

dbl

instrumentalness

A measure of the likelihood that the song does not contain vocals

⋮

lgl

liveness

A measure of the presence of a live audience in the recording

⋮

dbl

valence

A measure of the musical positiveness conveyed by the song

⋮

dbl

tempo

The tempo of the song in beats per minute

⋮

dbl

time_signature

The estimated overall time signature of the song

⋮