

# TikTok Popularity And Billboard Success

Lagunero, Raizel | Russell, Keyan | Tran, Alan

## About

This dataset is a list of popular songs on TikTok and their respective characteristics.

Alert Count	3
Completeness	0
Provenance	0
Collection	2
Selection Bias	2
Description	0
Composition	1
Population Bias	1

## Use Cases

Potential real-world applications of the dataset

1. What properties do popular songs on TikTok and top charting songs have?
2. What is the song attribute that tends to be the most effective in captivating its listeners?
3. Does an artist’s popularity influence what songs might become popular on TikTok?

Dataset Nutrition Label ★ ° ★ 90° ★ ° ★

## TikTok Popularity And Billboard Success

---

Alerts ⚠ ⚠ ⚠

### Collection: Selection Bias

- This dataset only takes songs that appear on Spotify *and* TikTok, so songs that are exclusive to Youtube, Apple Music, or TikTok won't be available for our dataset.
- There is a high chance that songs that appeared on TikTok and are listed on Spotify that aren't listed in our dataset.

### Composition: Population Bias

- Assuming that Billboard is the only reliable way to know a song's popularity because this data would only really be pulled from the US since Billboard's data is based off of US listeners' data, so we can eliminate the notion that this represents global trends.

# TikTok Popularity And Billboard Success

---

## Dataset Information

This dataset represents the TikTok and Spotify communities. The average listener is also represented by Billboard's data collection, through radio listeners, as well as other digital and physical media sales.

The following questions in this section is derived from Data Nutrition Project <https://datanutrition.org/labels/nopv-nyc/#Alerts-title> and Datasheets for Datasets <https://arxiv.org/pdf/1803.09010.pdf>.

---

## Description

- 1. Is there an intended purpose for the dataset? What domain was it designed for?**  
Describing aspects of popular songs on the app (danceability, tempo, etc.)
  - 2. Are there tasks for which the dataset would be cautioned against being used? If so, please provide a description.**  
Anything other than musical/media analysis because it would be irrelevant
- 

## Composition

- 1. Does the dataset have a metadata repository or data dictionary? If yes, please provide the link, and if not, please explain what each field means.**  
<https://spotify-api.js.org/main/interface/AudioFeatures>
- 2. What data does each instance consist of? For example, does it consist of raw data (E.G., unprocessed text or images) or features? In either case, please provide a description.**  
Each instance contains the song title, the artist who sang it, and a variety of song features labeled on different scales (0-1 for danceability, N/A or 1-100 for peak position on the Billboard Chart, etc.)
- 3. How many instances are there in total (of each type, if appropriate)?**  
263 instances
- 4. Is any information missing from individual instances? If so, what is missing and why?**  
None, the only instance of missing information was being on the Billboard 100, being FALSE reads as N/A, so data might appear to be missing, but the song is not performed by an artist who is in the Top 100. The other N/A values relate to positions on the chart, so that just means that the song was never in the Top 100.

5. **Was any preprocessing/cleaning/labeling of the data done? If so, please provide a description.**

A dataset containing popular TikTok songs was merged with a dataset containing Billboard Top 100 songs. An additional column containing the overall popularity of the song was added.

6. **Does the dataset identify any subpopulations (E.G., by age, gender)? If yes, please indicate their distributions within the dataset where possible/available.**

No

7. **Is the dataset self-contained, or does it link to or otherwise rely on external resources?**

No

---

## Collection

1. **Over what timeframe was the data collected? If the data continues to be updated, please indicate.**

Jan 2021 - Feb 2022

2. **What mechanisms or procedures were used to collect the data?**

Using Spotify's API and Billboard's Top 100 Charts

3. **If individuals' data is included in this dataset, did those individuals consent to the collection and use of their data? If so, please describe the consent procedure.**

N/A

4. **If individuals' data is included in this dataset, was this data altered to ensure higher levels of privacy? If yes, please describe any privacy procedures followed with regards to this dataset privacy procedures followed with regards to this dataset (anonymization efforts, privacy protocols, suppression techniques, etc).**

N/A

5. **Has this data been reviewed for quality?**

No

6. **Were any ethical review processes conducted (e.g., by an institutional review board)? If so, provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.**

No

7. **Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals race or ethnic origins, sexual orientations, or locations; financial or health data; biometric or genetic data; forms of government identification)?**

No