

DataHack 2021 Challenge

The Challenge

In addition to being a team of skilled data scientists, your DataHack 2021 group is also a band of hobbyist musicians that publish music. Your team has access to streaming/music reviews data from Spotify and Pitchfork.com, respectively, and your job is to leverage this data to help propel your amateur band to stardom! In other words, you want to think about the following question:

What makes a song successful?

You define success. It can be a KPI (# streams, critic reviews, etc.), or it can be something more qualitative. This problem can be tackled in a variety of ways, such as:

- Developing a business plan that outlines how your amateur music group will achieve success.
- Creating an ML model to predict the success of a song and analyzing the features which contribute to that success.
- Building a compelling dashboard/visualization in the context of successful music.

We encourage creativity! Feel free to tackle the challenge in any way as long as it focuses on the key question.

Competition Deliverables & Evaluation

By the end of the competition, your team is expected to deliver **(1)** a concise 5-minute (or less) PowerPoint presentation about your idea and key findings, and **(2)** any code that you write/use. Judges will evaluate your work via 3 categories:

1. **Creativity:** How novel are your ideas? Is your submission exciting?
2. **Technical Mastery:** Is your proposal backed up by mathematical evidence? Did you use any interesting technologies?
3. **Quality:** How effective is your presentation? How polished is your code? How well can you sell your idea?

(continued on next page)

Dataset Details

The dataset includes these features for you to work with.

Feature	Datatype	Description (if needed)
name	String	
album	String	
artist	String	
auditory	Float	
beats_per_measure	Integer	
beats_per_min	Float	
concert_probability	Float	Probability that song was recorded live
critic	String	
critic_rating	Integer	Higher rating = better review
danceability	Float	
hotness	Float	Time-sensitive popularity metric
hype	Float	
instrumentalness	Float	
length_minutes	Float	
lyricism	Integer	
major/minor	String	
nplays	Integer	Number of Spotify plays
positivity	Float	
review	String	
reviewer_type	String	
song_debut	String	Mostly in yyyy-mm-dd but precision may vary
styles	String	Comma separated list of styles
tone	String	
volume	Integer	
vulgar	String	

IMPORTANT: Pitchfork.com reviews are written on a *by-album* basis, i.e. each review corresponds to an entire *album*. Therefore, each song in an album will have the same pitchfork review/score.