

# Data Dictionary

There are 20 columns in this dataset and their description are as follows.

- Track id: unique identifier for each track used by Spotify (randomly generated alphanumeric string)
- Track name: track name
- Album id
- Playlist id
- Playlist name
- popularity: song popularity score as of March 2021 on a normalized scale [0-100] where 100 is the most popular
- duration\_ms: duration of track in milliseconds
- explicit: true or false if the song contains explicit content.
- artists: name of the main artist
- id\_artists: unique identifier for each artist used by Spotify
- release\_date: when the album was released (date format: yyyy/mm/dd)
- danceability: describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.
- energy: measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy.
- key: The estimated overall key of the track. Integers map to pitches using standard Pitch Class notation. E.g. 0 = C, 1 = C $\sharp$  / D $\flat$ , 2 = D, and so on. If no key was detected, the value is set to -1.
- loudness: The overall loudness of a track in decibels (dB). Values typical range between -60 and 0 db.
- mode: Mode indicates the modality (major=1 or minor=0) of a track, the type of scale from which its melodic content is derived.

- **speechiness**: measures from 0.0 to 1.0 and detects the presence of spoken words in a track. If the speechiness of a song is above 0.66, it is probably made of spoken words, a score between 0.33 and 0.66 is a song that may contain both music and words, and a score below 0.33 means the song does not have any speech.
- **acousticness**: confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic
- **instrumentalness**: measure from 0.0 to 1.0 and represents the amount of vocals in the song. The closer it is to 1.0, the more instrumental the song is.
- **liveness**: likelihood measure from 0.0 to 1.0 and indicates the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live.
- **valence**: A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track.
- **tempo**: The overall estimated tempo of a track in beats per minute (BPM)
- **time\_signature**: An estimated overall time signature of a track. The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure).
- **Popularity\_rate**: four-categorical variables depicting the popularity of the song in term of: low /mid-low/mid-high/high
- **PITCH**: 0/1 variables showing the PITCH/ NO PITCH of the song in different playlist