

## Informazioni sul dataset

### Contesto

Questo dataset contiene statistiche audio delle 2000 tracce più ascoltate su Spotify dal 2000 al 2019. I dati contengono 18 colonne, ciascuna delle quali descrive la traccia e le sue caratteristiche.

### Contenuto

- **artist:** nome dell'artista.
- **song:** nome della traccia.
- **duration\_ms:** durata della traccia in millisecondi.
- **explicit:** il testo o il contenuto di una canzone o di un video musicale contiene uno o più criteri che potrebbero essere considerati offensivi o inadatti ai bambini.
- **year:** anno di uscita della traccia.
- **popularity:** maggiore è il valore, più popolare è la canzone.
- **danceability:** la ballabilità descrive quanto una traccia sia adatta al ballo in base a una combinazione di elementi musicali, tra cui tempo, stabilità del ritmo, intensità del beat e regolarità generale. Un valore pari a 0,0 indica la minima ballabilità e 1,0 indica la massima ballabilità.
- **energy:** l'energia è una misura da 0,0 a 1,0 e rappresenta una misura percettiva di intensità e attività.
- **key:** la tonalità della traccia. Gli interi vengono mappati alle altezze utilizzando la notazione standard Pitch Class. Ad esempio, 0 = Do, 1 = Do#/Reb, 2 = Re e così via. Se non è stata rilevata alcuna tonalità, il valore è -1.
- **loudness:** il volume complessivo di una traccia in decibel (dB). I valori di volume sono calcolati come media sull'intera traccia e sono utili per confrontare il volume relativo delle tracce. Il volume è la qualità di un suono che è il principale correlato psicologico della forza fisica (ampiezza). I valori variano in genere tra -60 e 0 dB.
- **mode:** il modo indica la modalità (maggiore o minore) di una traccia, il tipo di scala da cui deriva il suo contenuto melodico. Maggiore è rappresentato da 1 e minore da 0.
- **speechiness:** il parlato rileva la presenza di parole parlate in una traccia. Più la registrazione è esclusivamente di tipo parlato (ad esempio, talk show, audiolibro, poesia), più il valore dell'attributo si avvicina a 1,0. Valori superiori a 0,66 descrivono tracce che sono probabilmente composte interamente da parlato. Valori compresi tra 0,33 e 0,66 descrivono tracce che possono contenere sia musica che parlato, sia in

sezioni che stratificati, inclusi casi come la musica rap. Valori inferiori a 0,33 rappresentano molto probabilmente musica e altre tracce non di tipo parlato.

- **acousticness**: una misura di confidenza da 0,0 a 1,0 sulla natura acustica della traccia. 1,0 rappresenta un'elevata confidenza che la traccia sia acustica.

- **instrumentalness**: prevede se una traccia non contiene parti vocali. I suoni "Ooh" e "aah" sono considerati strumentali in questo contesto. Le tracce rap o di parlato sono chiaramente "vocali". Più il valore di strumentalità è vicino a 1,0, maggiore è la probabilità che la traccia non contenga contenuti vocali. I valori superiori a 0,5 rappresentano tracce strumentali, ma l'affidabilità è maggiore man mano che il valore si avvicina a 1,0.

- **liveness**: rileva la presenza di un pubblico nella registrazione. Valori di liveness più elevati rappresentano una maggiore probabilità che la traccia sia stata eseguita dal vivo. Un valore superiore a 0,8 fornisce un'alta probabilità che la traccia sia dal vivo.

- **valence**: una misura da 0,0 a 1,0 che descrive la positività musicale trasmessa da una traccia. Le tracce con valenza elevata suonano più positive (ad esempio, felici, allegre, euforiche), mentre le tracce con valenza bassa suonano più negative (ad esempio, tristi, depresse, arrabbiate).

- **tempo**: il tempo complessivo stimato di una traccia in battiti al minuto (BPM). Nella terminologia musicale, il tempo è la velocità o il ritmo di un dato brano e deriva direttamente dalla durata media dei battiti.

- **genre**: genere della traccia.