Dataset Presentation

Spotify 1.2M+ Songs • *Mazza Luca, Hoxha Roeld, Masciocchi Andrea*

Spotify 1.2M+ Songs

What does the dataset represent?

- 🎶 **Songs** - Spotify's dataset of 1.2M+ entries (1'204'025)
- 📏 **Row** - Each row represents a song
- 🎤 **Columns** - Each column is a characteristic of the song (24 columns)
- ⏱️ **Time Span**: 1908 - 2020 (Publication Date)
- 🌍 **Worldwide** - Songs from all over the world

Interesting Columns

Some interesting columns to analyze

Song Name	Name of the song
Danceability `float(0,1)` **Energy** `float(0,1)`	Interesting to compare with eachother
Loudness `dB` **Speechiness** `float(0,1)` **Duration** `ms`	Interesting the change over the years
Publication Date `DateTime`	Most productive parts of the years

Spotify computes danceability, energy, … with ML algorithms

Why?

Why is this dataset interesting (to us)?

- Spotify is an every-day use for most of us
- Data Credibility
- Songs from all over the world
- Changes over time on music

Thank you!

Dataset · GitHub