Spotify Data Analysis

Do you have a top 100 song?

Can we predict a top 100 Spotify song?

1. Collect the Data

- ☐ Spotify API
- ☐ Spotify Charts
- ☐ Hot Songs of the Decade

2. Build a Model

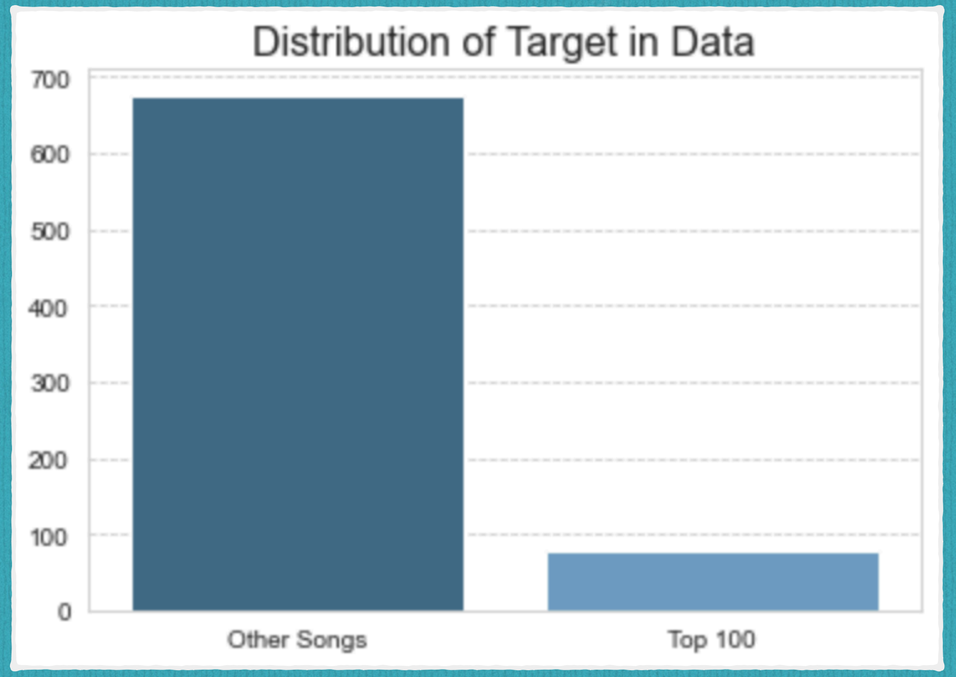
- ☐ Fix Class Imbalance
- ☐ Feature Engineering

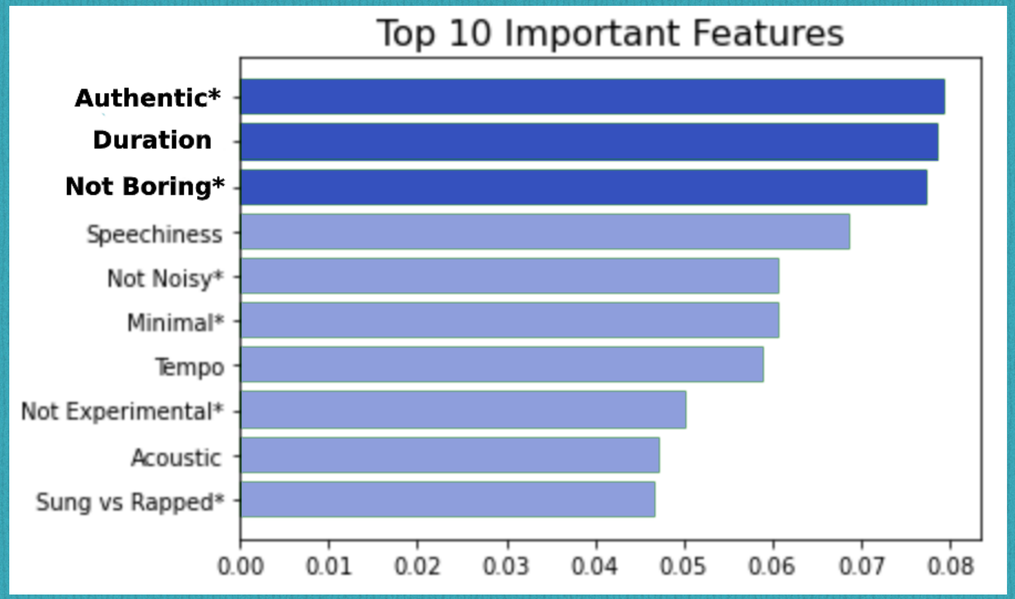
3. Assess the Model

- ☐ Most Influential Features
- ☐ Online Predictor

Building a Model

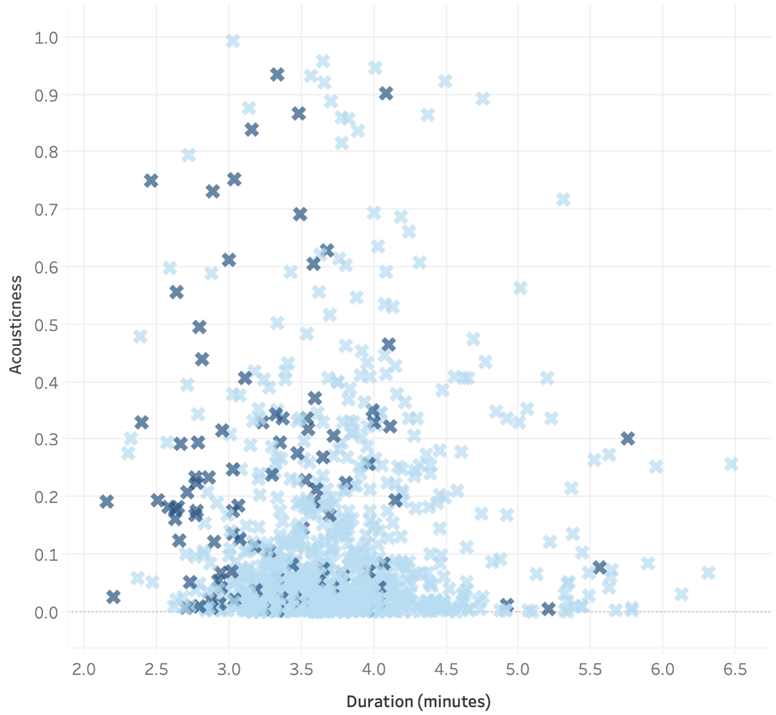
- Challenge 1: Fixing Imbalance
  - SMOTE 2:1
  - Class weights 2:1
  - Threshold Tuning (.485)
- Challenge 2: Capturing More
  - Feature Engineering*



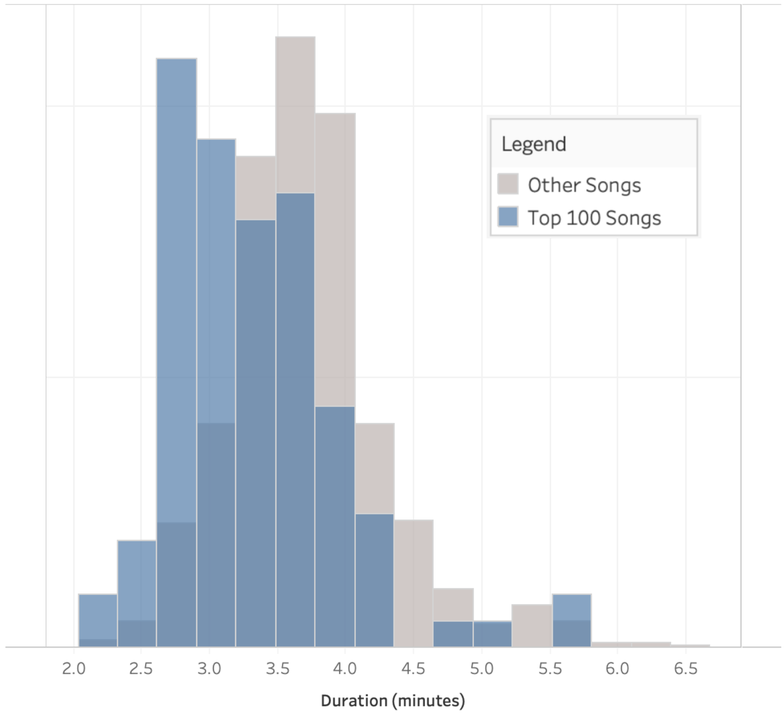


Top 3 Feature Conclusions

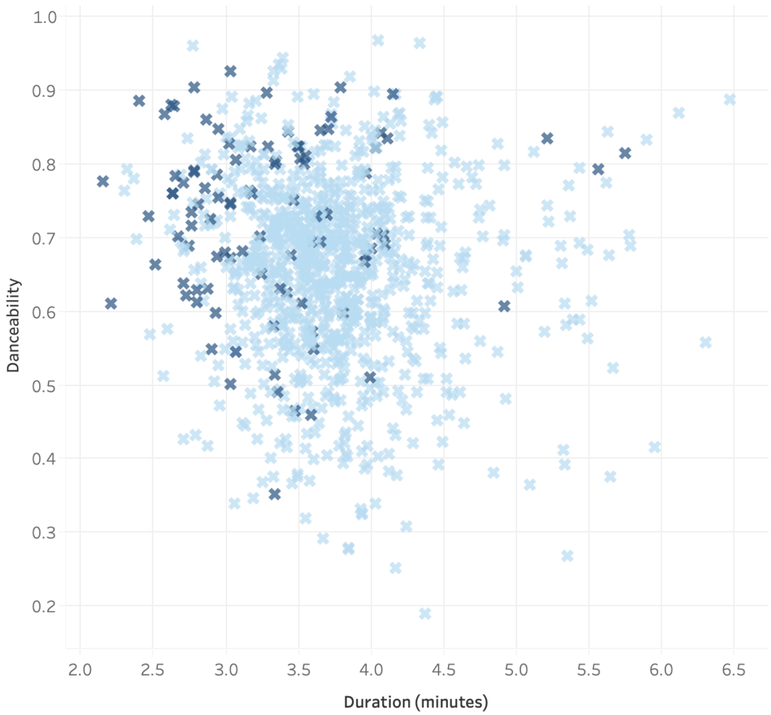
Should be "Authentic" / "Singer-Songwriter"



Shorter Songs are Better



Quick Dance Songs > Long and Drawn Out

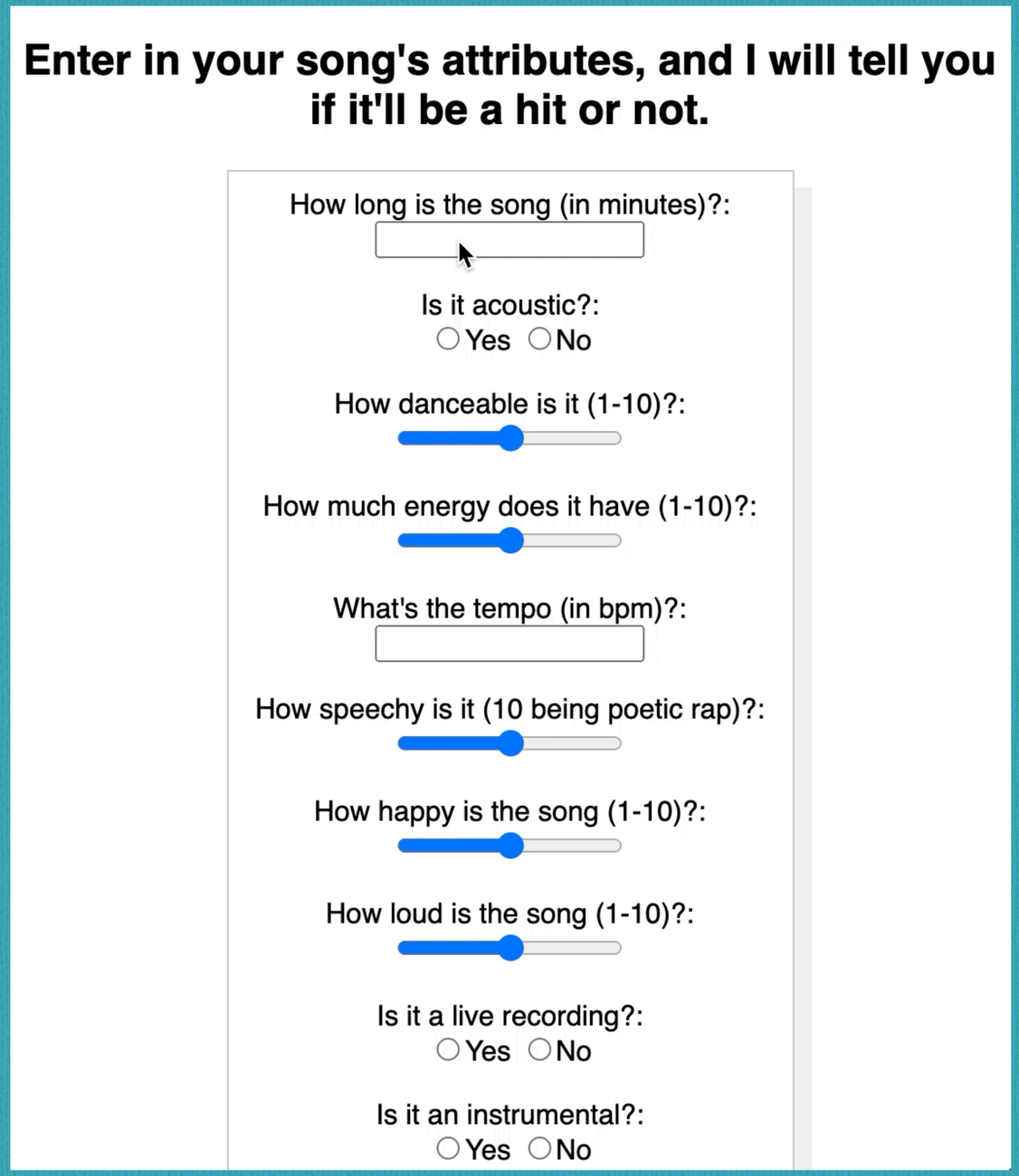


2/3 Engineered

Using the Model

Final Scores

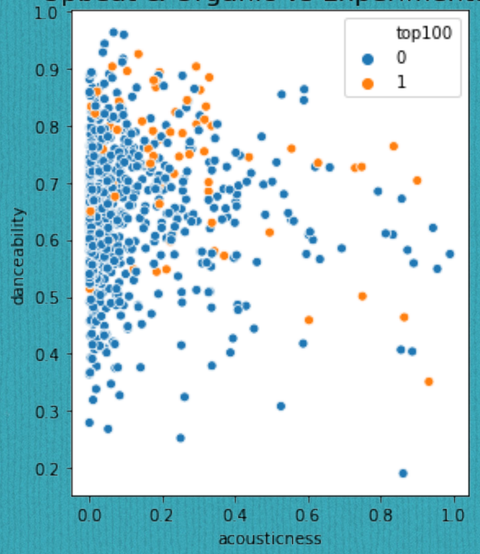
- ROC/AUC = .764
- F-Beta = .51
- Precision = .714



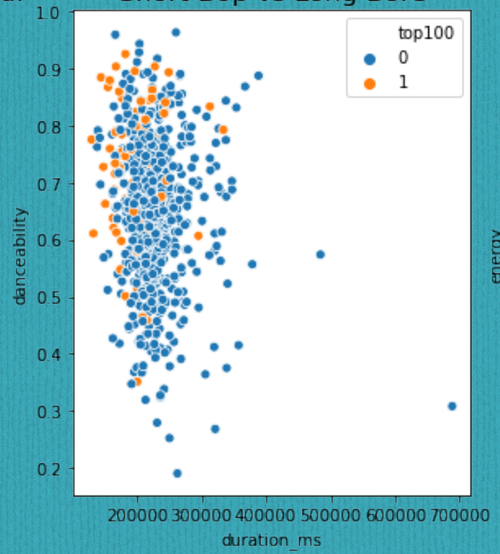
Questions?

Appendix - Engineered Features

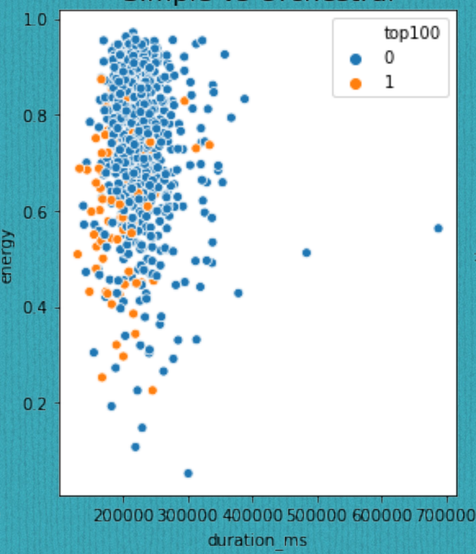
Upbeat & Organic vs Experimental



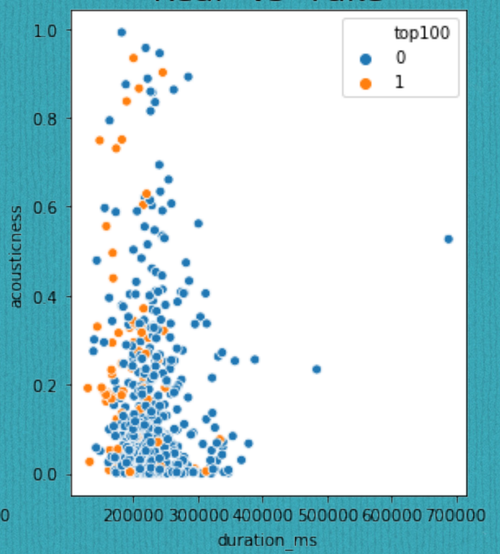
Short Bop vs Long Bore



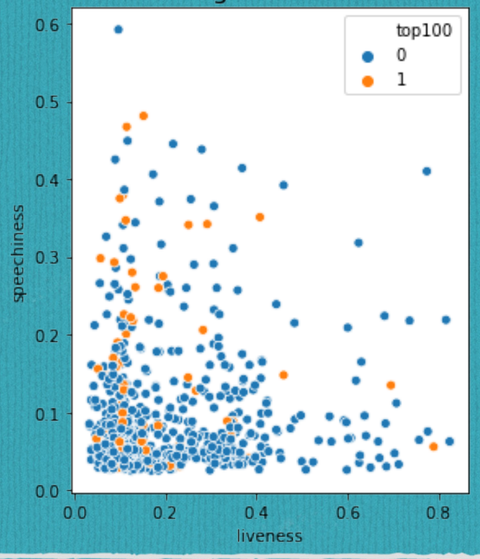
Simple vs Orchestral



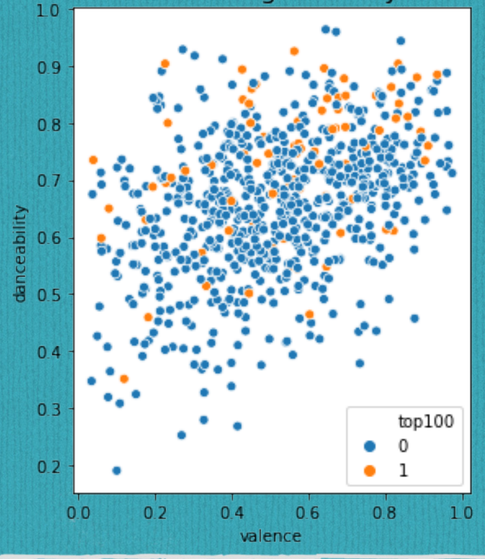
"Real" vs "Fake"



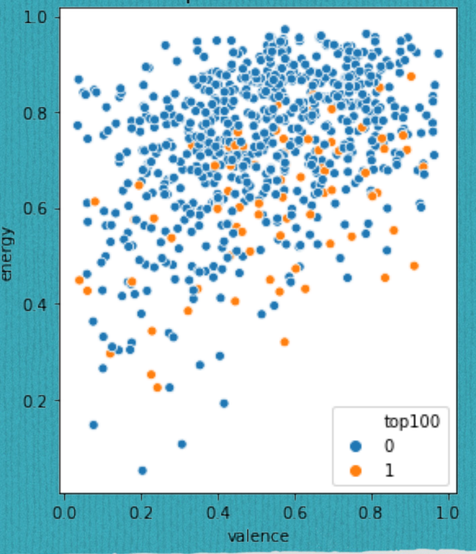
Song vs Podcast



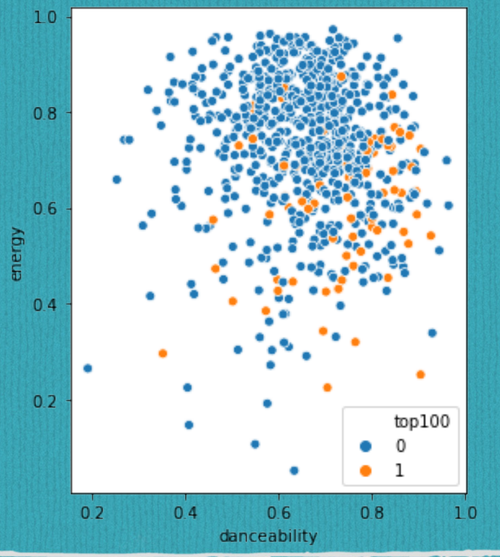
Club Song vs Corny



Pop vs Screamo

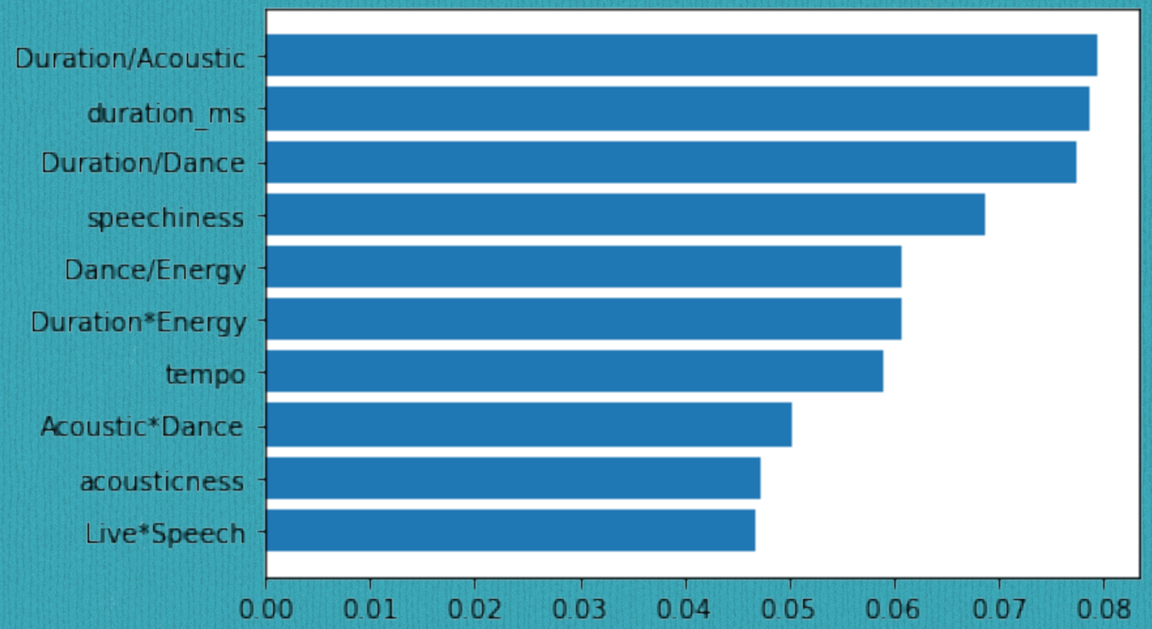


Music vs Noise



Appendix

Some Clarification



Valence - Hapiness

Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).

Energy

Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.

Danceability

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.