# SONG CLUSTERING

BY TANNER KOSCINSKI

# ATTAINING THE DATASET

- Spotify Web API
- Spotipy Python Library
- Search artists, albums, and playlists
- Get tracks from albums and playlists
- Get Spotify’s audio analysis for each track

# EXPLORING THE DATASET

- *"Timbre* is the quality of a musical note or sound that distinguishes different types of musical instruments, or voices. It is a complex notion also referred to as sound color, texture, or tone quality, and is derived from the shape of a segment's spectro-temporal surface, independently of pitch and loudness. The timbre feature is a vector that includes 12 unbounded values roughly centered around 0. Those values are high level abstractions of the spectral surface, ordered by degree of importance."
  - https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-analysis/
- Each song has hundreds of segments, and each segment has its own timbre vector

# CLUSTERING THE DATASET

- Cluster songs together based on the timbre sounds within the songs
- Practical use: sort large playlists into multiple smaller playlists
- Data product: website that takes a playlist as input and outputs multiple smaller playlists containing similar songs from the original playlist