

Group members: Emily Armstrong, Ayush Gupta, Sarika Gandigude
Project: Predicting the popularity of a song using metrics from Spotify

Dataset Description

The dataset was downloaded through Kaggle and is made up of 30,000 songs. The data was originally collected from Spotify.

The variables are:

Track ID

Track Name

Track Artist

Track Popularity

Track Album ID

Track Album Name

Track Album Release Date

Playlist Name

Playlist ID

Playlist Genre

Playlist Subgenre

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. A value of 0.0 is least danceable and 1.0 is most danceable.

Energy - Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. 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.

Key - The estimated overall key of the track. Integers map to pitches using standard Pitch Class notation . E.g. 0 = C, 1 = C#/D ♭ , 2 = D, and so on. If no key was detected, the value is -1.

Loudness - The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typical range between -60 and 0 db.

Mode - Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0.

Speechiness - Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.

Acousticness - A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.

Instrumentalness - Predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal". The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.

Liveness - Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.

Valence - A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. 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).

Tempo - The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.

Duration - measured in milliseconds

Project Description

What makes a song popular? With the right information, can the popularity of a song be predicted? That is what we would like to uncover. The data we have provides a well-rounded look at various aspects of music, giving us tangible data to try and make predictions about something typically seen as subjective (the popularity of a song). Will we observe certain keys more often than others? Is there a relationship between the popularity of a song and its valence? Or the key? We will dig deeper into the nuances of a track and use these nuances to make predictions about the popularity of songs. The era of streaming has had an impact on the length of songs, will we see this reflected in the popularity of songs over the time series of the data?