**Dataset “Spotify and Youtube”**

**Посилання на датасет:** [**\*Тик\***](https://www.kaggle.com/datasets/salvatorerastelli/spotify-and-youtube/data)

## 

## **About Dataset**

Dataset of songs of various artist in the world and for each song is present:

* Several statistics of the music version on spotify, including the number of streams;
* Number of views of the official music video of the song on youtube.

# **Content**

It includes 26 variables for each of the songs collected from spotify. These variables are briefly described next:

* **Track**: name of the song, as visible on the Spotify platform.
* **Artist**: name of the artist.
* **Url\_spotify**: the Url of the artist.
* **Album**: the album in wich the song is contained on Spotify.
* **Album\_type**: indicates if the song is relesead on Spotify as a single or contained in an album.
* **Uri**: a spotify link used to find the song through the API.
* **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**: 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 key the track is in. 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 typically range between -60 and 0 db.
* **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\_ms**: the duration of the track in milliseconds.
* **Stream**: number of streams of the song on Spotify.
* **Url\_youtube**: url of the video linked to the song on Youtube, if it have any.
* **Title**: title of the videoclip on youtube.
* **Channel**: name of the channel that have published the video.
* **Views**: number of views.
* **Likes**: number of likes.
* **Comments**: number of comments.
* **Description**: description of the video on Youtube.
* **Licensed**: Indicates whether the video represents licensed content, which means that the content was uploaded to a channel linked to a YouTube content partner and then claimed by that partner.
* **official\_video**: boolean value that indicates if the video found is the official video of the song.