Data Mining Weka Spotify Franco Baez.

In this project, I filtered the Spotify CSV dataset by selecting arbitrary quantities and relevant data for my research purposes. I aimed to identify common characteristics among musical hits spanning various genres and time periods. Utilizing Weka, I employed the J48 decision tree algorithm to analyze the dataset and uncover underlying patterns in the data

CSV Source: <https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023?resource=download>

This dataset contains a comprehensive list of the most famous songs of 2023 as listed on Spotify. The dataset offers a wealth of features beyond what is typically available in similar datasets. It provides insights into each song's attributes, popularity, and presence on various music platforms. The dataset includes information such as track name, artist(s) name, release date, Spotify playlists and charts, streaming statistics, Apple Music presence, Deezer presence, Shazam charts, and various audio features.

The dataset has songs from 1930-2023.

***Definition of the features on filtered dataset:***

* **Artist\_count:** Sometimes songs are a team effort, and this tells us how many people worked on a song. It could be one person or a group of artists.
* **Streams**: It’s like how many times people press the play button on Spotify. The more streams, the more people like the song.
* **Bpm:** Think of this as the heartbeat of the song. It’s how fast or slow it goes.
* **Danceability\_%:** This shows how much your feet will want to dance to the song. It’s like a dancing score.
* **Valence\_%:** This tells us if the song feels happy or not. The higher, the happier.
* **Energy\_%:** It’s like the song’s battery level. High energy means the song is super active.
* **Acousticness\_%:** Imagine this as the “acoustic guitar” level. High acousticness means the song sounds natural.
* **Liveness\_%:** It’s like having a live concert in your headphones. High liveness means you can feel the energy of a live performance.
* **Speechiness\_%:** This is about talking in the song. High ‘speechiness’ means there are more spoken words.

BPM:

* High BPM (>140)
* Medium BPM (>=90=<140)
* Low BPM (<90)

Danceability:

* High Danceability (>70)
* Medium Danceability (>=50<=70)
* Low Danceability (<50)

Valence:

* High Valence (>70)
* Medium Valence (>=50<=70)
* Low Valence (<50)

Energy:

* High Energy (>70)
* Medium Energy (>=50<=70)
* Low Energy (<50)

Acousticness:

* High Acousticness (>70)
* Medium Acousticness (>=50<=70)
* Low Acousticness (<50)

Liveness:

* High Liveness (>70)
* Medium Liveness (>=50<=70)
* Low Liveness (<50)

Liveness:

* High Speechiness (>55)
* Medium Liveness (>=35<=55)
* Low Liveness (<35)

Hits:

* Big Hit (>1 Billion Streams)
* Medium Hit (>=500 Million Streams <= 1 Billion Streams)
* Small Hit (<500 Million Streams)