CSE332 Introduction to Visualization

Seungjun Chae 112364236

**Lab Project #1**

1. Spreadsheet

[GitHub](https://github.com/sjchae416/CSE332-lab-project-1.git) link (hyperlink) to the dataset in spreadsheet and future project.

1. Attributes

There are 2 categorical and 8 numerical attributes of 500 data in total: track name, artist names, peak rank, weeks on chart, danceability, energy, loudness, speechless, instrumentalness, tempo.

Categorical (2):

* **Track name (500 songs)**: Name of track
* **Artist names (349 artists)**: Names of all artists who participated in the song

Numerical (8):

* **Peak rank (1 ~ 198 rank)**: Highest rank the song achieved on Spotify Charts
* **Weeks on chart (1 ~ 269 weeks)**: Number of weeks the song was on Spotify Charts
* **Danceability (0.308 ~ 0.956)**: 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 (0.133 ~ 0.962)**: 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.
* **Loudness(-16.169 ~ -1.322 dB)**: 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 (0.0232 ~ 0.611)**: 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.
* **Instrumentalness (0 ~ 0.908)**: 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.
* **Tempo (66.165 ~ 205.863 BPM)**: 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.

1. Dataset

* Name of the source with URL(hyperlink): [Spotify top chart songs 2022](https://www.kaggle.com/datasets/sveta151/spotify-top-chart-songs-2022)
* There are 500 data points in total.
* There are 10 columns and 501 rows (including the first line of header categories)

1. Justification

I chose this dataset because I was curious whether there is a relationship between audio features and a song being on the chart.

This dataset includes the name of the song and a singer and its peak rank and the duration of how long it was on chart. Most interestingly, it contains 6 numerical data on audio features so that I can conduct a project to figure out which one(s) of those could affect the rank or the duration on the chart. There are five hypotheses on this dataset:

1. Danceability audio feature of the track will affect the peak rank and the weeks on chart.
2. Energy audio feature of the track will affect the peak rank and the weeks on chart.
3. Speechiness audio feature of the track will affect the peak rank and the weeks on chart.
4. Instrumentalness audio feature of the track will affect the peak rank and the weeks on chart.
5. Tempo audio feature of the track will affect the peak rank and the weeks on chart.

When the conclusion is provided after the project, it will allow musicians to take notes on which audio features to put to produce a song with higher chance to stay on chart for a long period of time.