SPOTIFY SONGS ANALYSIS DOCUMENTATION

The dataset for this project was downloaded from Kaggle using this link: [spotify\_songs](https://www.kaggle.com/datasets/sujaykapadnis/spotify-songs/data). The dataset was chosen because:

1. Easy to understand. It’s easily comprehensible.
2. Relatable. Music is relatable to most people.
3. Beginner friendly. The dataset is perfect for person seeking to build their first data analysis project.

Data Cleaning & Transformations

1. Extraction of year from release date.
2. Total songs measure to get the distinct count of songs.
3. Convert songs durations from milliseconds to minutes by creating a separate column
4. Convert danceability, energy, speechiness, acousticness, liveness and valence fields to percentages.

Problem Statement

1. ~~Count of unique songs per genre to understand diversity~~
2. ~~How many total unique songs exist to highlight comprehensive overview of the data~~
3. ~~Top 5 years with most song releases to show music trends over time~~
4. ~~Top 10 artists with most albums to understand highlight prolific artists~~
5. Which are the top 10 albums with highest average track popularity by genre to understand diversity of albums
6. Which genres are having most songs included in playlists to showcase the extensiveness of tracks?
7. What is the average song duration for the different genres to highlight average genre length?
8. Rankings of danceability of genres to highlight preference of listeners.
9. Comparison of loudness and danceability to highlight if the relationship between the two.
10. ~~Average song duration per genre over the years to showcase music length trends over the years.~~