Jack Ambery

Professor Casement

STAT 2200 Statistical Computing

21 March 2024

<center>Final Project Proposal</center>

Dataset URL: https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023

This dataset has logged the top 953 most popular songs on Spotify in 2023. The dataset has variables like song title, artists, release year, streams, and even ratings for danceability, energy, liveliness, and "speechiness." I assume all of this data was collected in one moment at some point after 2023. Kaggle says it was last updated 7 months ago, so maybe it was collected around August 2023.

Variables and descriptions provided from Kaggle:

- **track_name**: *Name of the song*

- **artist(s)_name**: *Name of the artist(s) of the song*

- **artist_count**: *Number of artists contributing to the song*

- **released_year**: *Year when the song was released*

- **released_month**: *Month when the song was released*

- **released_day**: *Day of the month when the song was released*

- **in_spotify_playlists**: *Number of Spotify playlists the song is included in*

- **in_spotify_charts**: *Presence and rank of the song on Spotify charts*

- **streams**: *Total number of streams on Spotify*

- **in_apple_playlists**: *Number of Apple Music playlists the song is included in*

- **in_apple_charts**: *Presence and rank of the song on Apple Music charts*

- **in_deezer_playlists**: *Number of Deezer playlists the song is included in*

- **in_deezer_charts**: *Presence and rank of the song on Deezer charts*

- **in_shazam_charts**: *Presence and rank of the song on Shazam charts*

- **bpm**: *Beats per minute, a measure of song tempo*

- **key**: *Key of the song*

- **mode**: *Mode of the song (major or minor)*

- **danceability_%**: *Percentage indicating how suitable the song is for dancing*

- **valence_%**: *Positivity of the song's musical content*

- **energy_%**: *Perceived energy level of the song*

- **acousticness_%**: *Amount of acoustic sound in the song*

- **instrumentalness_%**: *Amount of instrumental content in the song*

- **liveness_%**: *Presence of live performance elements*

- **speechiness_%**: *Number of spoken words in the song*

These are the variables I will look at in my project:

Variable Name	Quantitative or Categorical?	What it measures	Units
streams	Quantitative	Total number of streams on Spotify	Streams – just an integer

bpm	Quantitative	Beats per minute, a measure of song tempo	Beats per minute – also just an integer
key	Categorical	Key of the song	Musical Key (A, B, C, D, E, F, G)