Appendix 1: Dictionary

The columns in our dataset with their definitions are provided below (Spotify AB, 2020) (Ansari, 2020):

- track: The Name of the track.

- artist: The Name of the Artist.

- uri: The resource identifier for the track.

- danceability: 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.

    Examples:

    o   0.1: Through the Fire and Flames- Dragon Force

    o   1.0: Around the world- Daft Punk

- energy: 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. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.

    Examples:

    o   0.0: Stardust- Frank Sinatra

    o   1.0: Nightrain- Guns N' Roses

- key: The key the track is in. Integers map to pitches using standard Pitch Class notation.

Integer	Key
0	C
1	C#
2	D
3	Eb
4	E
5	F
6	F#
7	G
8	G#
9	A
10	Bb
11	B

- 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 typical range between -60 and 0 db.

- mode: The modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0.

- speechiness: 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. The distribution of values for this feature look like this

- 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).

  Examples:
  - 0.0: Hypnotize - System of a Down
  - 1.0: Low Rider - War

- 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.

- time_signature: An estimated overall time signature of a track. The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure).

- chorus_hit: This the the author's best estimate of when the chorus would start for the track. Its the timestamp of the start of the third section of the track. This feature was extracted from the data received by the API call for Audio Analysis of that particular track.

- sections: The number of sections the particular track has. This feature was extracted from the data received by the API call for Audio Analysis of that particular track.

- target: The target variable for the track. It can be either '0' or '1'. '1' implies that this song has featured in the weekly list (Issued by Billboards) of Hot-100 tracks in that decade at least once and is therefore a 'hit'. '0' Implies that the track is a 'flop'.
  - The author's condition of a track being 'flop' is as follows:
    - The track must not appear in the 'hit' list of that decade.
    - The track's artist must not appear in the 'hit' list of that decade.
    - The track must belong to a genre that could be considered non-mainstream and / or avant-garde.
    - The track's genre must not have a song in the 'hit' list.
    - The track must have 'US' as one of its markets.

- track id: The Spotify track_id of the music track

- artist id: The Spotify artist_id of the track artist

- genres_1 to genres_4: The first 4 subgenres of the artist listed on Spotify

- genres: The main genre categorized based on the subgenre

- Decade: The decade of when this track achieved commercial success

- key_name: The key of the track converted into English

- mode_name: The mode of the track converted into English i.e. 0: Minor, 1: Major

- key_full: The name of the key along with the mode that the track follows. Eg: D Minor

- hit_flop: target converted from 1/0 into Hit/Flop.