Spotify hit songs data description

|  |  |  |
| --- | --- | --- |
| FEATURE | FEATURE DESCRIPTION | NEEDED? |
| explicit | Whether or not the track has explicit lyrics |  |
| track\_number | track’s index relative to its album | helps. Lower the number, the more listeners that song has |
| num\_artists | Number of artists who perform the song. | ??? – not needed |
| num\_available\_markets | The markets in which the album is available: [ISO 3166-1 alpha-2 country codes](http://en.wikipedia.org/wiki/ISO_3166-1_alpha-2). ***NOTE****: an album is considered available in a market when at least 1 of its tracks is available in that market.* | not needed  Not useful in determining. And this is not fixed. Can be changed anytime by the artist |
| release\_date | Release date of the song | only keep month, and not year or date – for prediction  keep the other 2 columns for EDA. |
| duration\_ms | The duration of the track in milliseconds |  |
| key | The key the track is in. Integers map to pitches using standard [Pitch Class notation](https://en.wikipedia.org/wiki/Pitch_class). E.g. 0 = C, 1 = C♯/D♭, 2 = D, and so on. If no key was detected, the value is -1 |  |
| mode | Mode indicates 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 |  |
| time\_signature | An estimated time signature. The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure). The time signature ranges from 3 to 7 indicating time signatures of "3/4", to "7/4".  Range: 3 – 7  Example: 4 |  |
| 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 |  |
| danceability | 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 | 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. |  |
| 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 |  |
| 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 typically range between -60 and 0 db. |  |
| speechiness | 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. |  |
| 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).  Range: 0 – 1  Example: 0.428 |  |
| 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.  Example: 118.211 |  |

SMART QUESTIONS:

1. **Cross-Genre Cluster Collaborations:** Analyze the impact of cross-genre collaborations on song success. Do songs that merge different musical styles tend to perform better?
2. **Artist Follower Count vs. Popularity:** Explore the relationship between the number of followers an artist has and the popularity of their collaborative songs. Do artists with more followers produce more popular songs?
3. **Valence and Success:** Can you analyze the sentiment of song lyrics (positive, negative, neutral) and its relationship with song popularity? Do songs with more positive lyrics tend to be more popular?
4. **Temporal Success Patterns:** Are there specific months or seasons when songs tend to be more successful? How does the release date (e.g., season, day of the week) impact song popularity?
5. **Influence of Explicit Content on Different Genres:** Does the impact of explicit content on song popularity vary across different genres? Are explicit lyrics more accepted in certain genres?

* Models – logistic regression, decision tree, xgboost+ TABLEAU

Reference

* https://www.billboard.com/music/music-news/the-importance-of-album-track-order-in-the-digital-age-1082814/