HitFinder Project Report

Impact of Audio-Related Features on Song Popularity

Objective

The objective of the HitFinder Python project is to determine the impact of various audio-related features on the popularity level of songs in Western popular music. This analysis aims to provide insight into which features contribute most significantly to a song's success.

Data for the project is based on Spotify's API [1], which provides access to the pertinent data for millions of tracks. The API is used directly, in addition to the Python package `spotipy` [2] which is useful for having already built out some of the custom functions needed to access the Spotify data.

Process

`1_get_data.ipynb`

Track data from Spotify's biggest artists of the past year and a half was used. The artists named on kworb.net [3] were searched for their artist ID with the API, which was then used to obtain the IDs of at most 20 of their albums (including albums, EPs, and singles), which was chunked into groups of 50 albums. Track data from the albums were then obtained. With just under 50,000 albums returned, album chunks were made since Spotify's API has a usage limit preventing copious amounts of function calling. Even then, a trimming of the albums to be used was necessary. Initial testing with public playlists is also shown.

`2_data_prep_and_modeling.ipynb`

Data preparation and modeling is performed on all data obtained in the previous code file. Data preparation steps include exploration (univariate analysis, outlier analysis) and cleaning (nulls, outliers, categorical feature encoding). Feature importance analysis is also done afterwards, where Shapley values [4], a metric for determining feature impacts upon a model's final prediction, are used. This portion works in tandem with the initial modeling process, where we see validation set predictions and metrics. (The modeling process will be expanded upon in the future.)

Conclusions

Analysis shows that audio features provided by Spotify's API have a relatively low impact on predicting a track's `popularity`. Both the Root Mean Squared Error (RMSE) and the mean prediction error on the validation set show only slight improvements as more features are added to the model. This suggests that while some features do influence popularity, their overall impact is limited.

Features and Impacts

14 distinct features were obtained with the objective of predicting the popularity of any given Spotify track, using thousands of tracks of training data. In the end, only eight features deemed worthwhile were investigated for their impact. Their definitions are copied from Spotify's API reference for tracks [5] and reference for audio features of tracks [6]. Feature impacts listed below are rated relative to one another.

1. **Explicit Content (`explicit`)**
   - **Definition**: Whether or not the track has explicit lyrics (`true` = yes it does; `false` = no it does not OR unknown). Converted to 1 for explicit label, 0 for no explicit label. This is not listed in the "audio feature" attributes in Spotify's API, but rather in the track attributes.
   - **Impact**: Very significant.
   - **Insight**: Tracks labeled as explicit tend to be more popular than those without the explicit label. This could be due to the appeal of edgier, more provocative content in popular music. For example, more topical or controversial music could tend to make more headlines, bringing more attention and therefore more clicks to a track.
2. **Duration (`duration_ms`)**
   - **Definition**: The duration of the track in milliseconds.
   - **Impact**: Significant.
   - **Insight**: Shorter songs are more popular, a trend that has been exacerbated by the influence of social media platforms like TikTok. These platforms favor shorter, more engaging content, leading to a preference for shorter tracks in contemporary popular music.
3. **Liveness (`liveness`)**
   - **Definition**: 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.
   - **Impact**: Moderate.
   - **Insight**: Songs that sound more like live performances tend to be less popular. This could be due to the decline in the popularity of live albums and the shift towards studio-produced tracks, as well as the availability of live performance videos on platforms like YouTube.
4. **Speechiness (`speechiness`)**
   - **Definition**: 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.
   - **Impact**: Moderate.
   - **Insight**: Tracks with higher “speechiness” (more spoken-word content), are generally less popular. This suggests that listeners prefer songs with more musical and melodic elements over those with a higher proportion of and/or focus on spoken words.
5. **Loudness (`loudness`)**
   - **Definition**: 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.
   - **Impact**: Moderate.
   - **Insight**: Louder tracks are more popular. This feature’s more extreme negative values indicate quieter tracks, which are less favored. The preference for louder music could be linked to its ability to capture attention and create a more engaging listening experience.
6. **Tempo (`tempo`)**
   - **Definition**: 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.
   - **Impact**: Low.
   - **Insight**: Slower tracks tend to be more popular. This might reflect a preference for more relaxed and mellow music, which can be more appealing for a wider audience.
7. **Danceability (`danceability`)**
   - **Definition**: 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.
   - **Impact**: Low.
   - **Insight**: More danceable tracks tend to have higher popularity values. This is likely influenced by social media trends, where danceable tracks often become viral due to choreographed dance routines.
8. **Valence (`valence`)**
   - **Definition**: 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).
   - **Impact**: Low.
   - **Insight**: Interestingly, tracks with more negative or sad tones are more popular. However, this feature has the least impact on the overall prediction, so this insight should be considered with caution.

Features Removed

1. **Energy (`energy`)**
   - **Definition**: 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.
   - **Reason for removal**: Too highly correlated to `loudness`. Inclusion of highly-correlated features can create model bias, so the least impactful of the two was removed.
2. **Key (`key`)**
   - **Definition**: The key the track is in. Integers map to pitches using standard Pitch Class notation [7]. E.g. 0 = C, 1 = C♯/D♭, 2 = D, and so on. If no key was detected, the value is -1.
   - **Reason for removal**: Low variation. A categorical feature, the key needed to be encoded numerically, which led to an error in building due to low variance in the encoded features. Regardless, the key had a very low impact even when included in the model.
3. **Mode (`mode`)**
   - **Definition**: 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.
   - **Reason for removal**: Low impact (lowest of all features).
4. **Acousticness (`acousticness`)**
   - **Definition**: A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
   - **Reason for removal**: Low impact.
5. **Instrumentalness (`instrumentalness`)**
   - **Definition**: 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.
   - **Reason for removal**: Low impact.
6. **Time Signature (`time_signature`)**
   - **Definition**: 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".
   - **Reason for removal**: Low variance. The vast majority of tracks sourced are in 4/4 time, meaning the pool of values outside of that signature was too low in count to determine a true impact. Even when grouping the other time signatures together, variance is still too low.

Summary

In summary, while certain audio-related features do influence the popularity of songs, their overall impact is relatively low. The most significant insights include the higher popularity of explicit content, shorter durations, and louder tracks. Social media trends, particularly those driven by platforms like TikTok, play a crucial role in shaping these preferences. Understanding these trends can help artists and producers tailor their music to better align with current listener preferences.

Next Steps

As the features used in this analysis were not of great use to predicting a track's popularity, obtaining different features would be a great next step. Spotify's API also contains more complex "audio analysis" features [8], which could provide better insight with features such as `timbre`. The developer API for last.fm [9], an app to track user listening habits, might be of use, although this will likely only give more popularity-related features rather than audio-related ones. Whether or not new features are acquired, feature engineering can also be performed to determine whether Spotify's features were used to their maximum capability.

Different model types and hyperparameter tuning can also be performed once more impactful features are discovered.

Sources

[1] https://developer.spotify.com/documentation/web-api

[2] https://spotipy.readthedocs.io/en/2.24.0/#api-reference

[3] https://kworb.net/spotify/listeners.html

[4] https://shap.readthedocs.io/en/latest/example_notebooks/overviews/An%20introduction%20to%20explainable%20AI%20with%20Shapley%20values.html

[5] https://developer.spotify.com/documentation/web-api/reference/get-track

[6] https://developer.spotify.com/documentation/web-api/reference/get-audio-features

[7] https://en.wikipedia.org/wiki/Pitch_class

[8] https://developer.spotify.com/documentation/web-api/reference/get-audio-analysis

[9] https://www.last.fm/api