Spotify Data Analysis

Arkin Satija

Overview

Dataset: https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset

What can we do with this?

- Using this data set we can look at over 90000 songs and their data and attempt to draw connections between each of the categories.
- With these connections drawn and looked at, we can make recommendations on how to make the next “most popular” song.

Understanding the Dataset

- **track_id**: The Spotify ID for the track
- **artists**: The artists' names who performed the track. If there is more than one artist, they are separated by a `;`
- **album_name**: The album name in which the track appears
- **track_name**: Name of the track
- **popularity**: **The popularity of a track is a value between 0 and 100, with 100 being the most popular**. The popularity is calculated by algorithm and is based, in the most part, on the total number of plays the track has had and how recent those plays are. Generally speaking, songs that are being played a lot now will have a higher popularity than songs that were played a lot in the past. Duplicate tracks (e.g. the same track from a single and an album) are rated independently. Artist and album popularity is derived mathematically from track popularity.
- **duration_ms**: The track length in milliseconds
- **explicit**: Whether or not the track has explicit lyrics (true = yes it does; false = no it does not OR unknown)
- **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
- **key**: The key the track is in. Integers map to pitches using standard Pitch Class notation. E.g. `0 = C`, `1 = C♯/D♭`, `2 = D`, and so on. If no key was detected, the value is -1
- **loudness**: The overall loudness of a track in decibels (dB)
- **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
- **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
- **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
- **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
- **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)
- **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
- **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`.
- **track_genre**: The genre in which the track belongs

Data

We are going to use the columns: ["track_id", "track_name", "artists", "popularity", "danceability", "energy", "tempo", "track_genre"] to create data that looks like this.

	track_id	track_name	artists	popularity	danceability	energy	tempo	track_genre
0	5SuOikwiRyPMVoIQDJUgSV	Comedy	Gen Hoshino	73	0.676	0.4610	87.917	acoustic
1	4qPNDBW1i3p13qLCt0Ki3A	Ghost - Acoustic	Ben Woodward	55	0.420	0.1660	77.489	acoustic
2	1iJBSr7s7jYXzM8EGcbK5b	To Begin Again	Ingrid Michaelson;ZAYN	57	0.438	0.3590	76.332	acoustic
3	6lfxq3CG4xtTiEg7opyCyx	Can't Help Falling In Love	Kina Grannis	71	0.266	0.0596	181.740	acoustic
4	5vjLSffimiIP26QG5WcN2K	Hold On	Chord Overstreet	82	0.618	0.4430	119.949	acoustic
...	...	...	...	...	...	...	...	...
113544	2C3TZjDRiAzdyViavDJ217	Sleep My Little Boy	Rainy Lullaby	21	0.172	0.2350	125.995	world-music
113545	1hIz5L4IB9hN3WRYPOCGPw	Water Into Light	Rainy Lullaby	22	0.174	0.1170	85.239	world-music
113546	6x8ZfSoqDjuNa5SVP5QjvX	Miss Perfumado	Cesária Evora	22	0.629	0.3290	132.378	world-music
113547	2e6sXL2bYv4bSz6VTdnfLs	Friends	Michael W. Smith	41	0.587	0.5060	135.960	world-music
113548	2hETkH7cOfqmz3LqZDHZf5	Barbincor	Cesária Evora	22	0.526	0.4870	79.198	world-music

113549 rows × 8 columns

Questions

What’s the most popular genre?

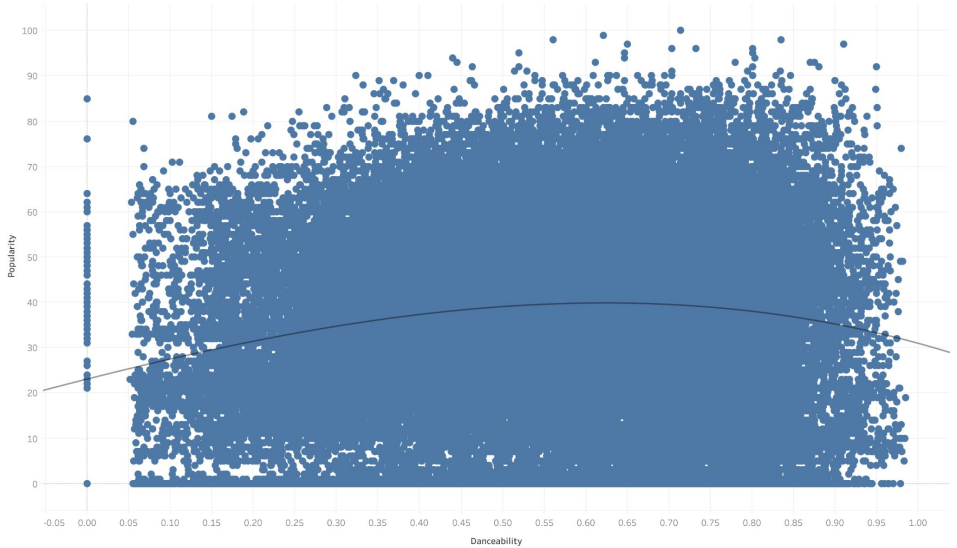
- Pop-film is the most popular genre

track_genre	popularity
pop-film	59.280280
k-pop	56.963928
chill	53.704705
sad	52.379000
grunge	49.582583

Questions

How does a song's danceability relate to the song's popularity?

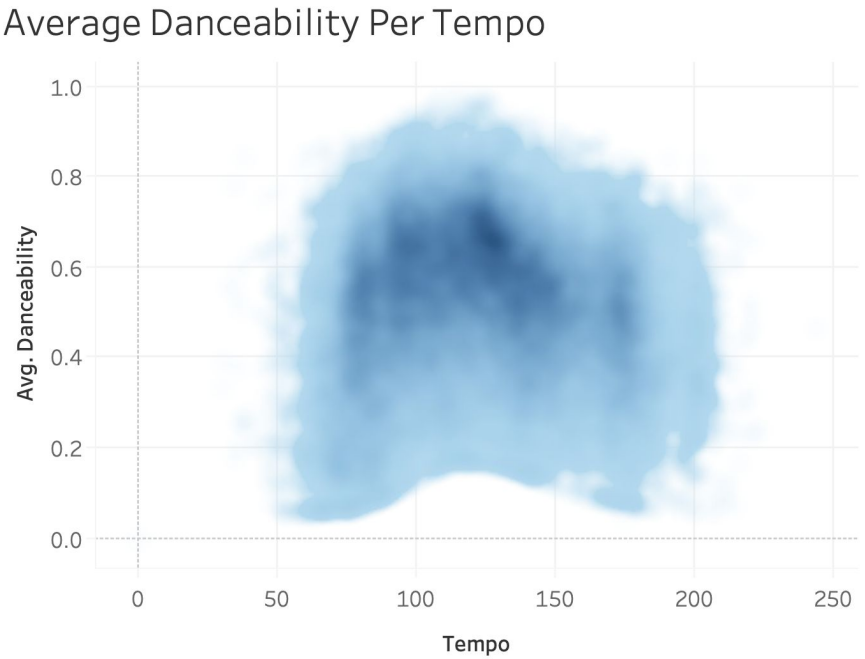
- We can see here that a slightly above halfway danceability is ideal for creating a popular song.



Questions

How does a song's tempo affect the danceability?

- Looking at this density plot, it looks like the most danceable tempo is around 125 BPM.



Recommendations

If you’re looking to make the next hit song, by spotify’s metrics, here’s what you can do:

- Make a movie featured pop song or a k-pop song.
- Make a song around a 0.6-0.65 danceability.
- Make a song with tempos between 100 BPM and 150 BPM
    - This is to make the song have a target danceability of 0.6 which was the most popular danceability of a song.

Next Steps

- Take a deeper look at how a song’s danceability relates to the popularity.
- Look at the statistics of only songs in the top 5 genres.
- Create a direct comparison between tempo and a song’s popularity

Thank you

My name is **Arkin Satija** and I am an aspiring Data Scientist.

You can reach out to me on linkedin:
https://www.linkedin.com/in/arkin-satija-b78446201