

Tracking Popularity of 2019 Music

Goal: Determine the popular songs of 2019 songs based on the elements acquired from their artists in 2018.

- We are searching for patterns in the audio features of the songs. Why do people stream these songs the most?
 - Analyze musical elements such as danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, duration and time and how these elements contribute to the popularity of music in 2019 - based upon its trends in 2018.
- <https://www.kaggle.com/nadintamer/top-spotify-tracks-of-2018>

General Project Description:

We acquired a dataset from Kaggle that accumulated songs top songs for 2018 and included their sound effects. We then used Exportify to get a list of songs made by these artists in 2019. We also used our own code to get the acoustic effects of these songs by generating our own CSV file using Spotify developer. We then analyzed the network of these artists and these song traits in order to create predictions for 2019.

To Do List:

1. Create a playlist of the songs used in the 2018 dataset to get a feel of their content
 - a. <https://github.com/mborgerson/spotify-playlist-from-csv>
2. Create a playlist of popular music in 2019 → convert into csv with exportify
 - a. <https://github.com/watsonbox/exportify>
3. Identify clusters of musical elements that contribute to popularity (2018 dataset)
4. Get audio features of 2019 popular music
 - a. <https://developer.spotify.com/documentation/web-api/reference/tracks/get-audio-features/>
5. Compare to 2019 popular music playlist (csv) → track networks of artists
6. Find correlations between:
 - a. Danceability and energy
 - b. Danceability and valence
 - c. Valence and tempo
 - d. Speechiness and tempo

Why was this dataset created?

- This dataset was created to analyze Spotify's top 100 tracks of the year 2018. The dataset includes Spotify's calculated features of the music: for example, qualities such as danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, duration and time

Who funded the creation of the dataset?

- This dataset was created using Spotify's web API and Spotipy library - Spotify created and calculated the audio features

What processes might have influenced what data was observed and recorded and what was not?

- The key influence for what songs were chosen is pure popularity (# of streams). As for the music elements examined, Spotify's features seek to interpret and analyze elements of the music that are responsible for certain activities (ex. dancing).

What preprocessing was done, and how did it come to be in the form that you are using?

- The preprocessing of the information was completed by Spotify, and was used internally for proprietary recommendation systems as well as popularity analytics.

If people are involved, were they aware of the data collection and if so, what purpose did they expect the data to be used for?

- No.

Interacting with our Data Set

This could be something as simple as descriptive statistics: how many rows, how many variables? Better would be some initial explorations.

- [Notebook]

Variables:

- *Id*: spotify URI of the song
- *Name*: name of the song
- *Artist*: artist(s) of the song
- *Danceability*: how suitable a track is for dancing based on tempo, rhythm, beat strength, regularity, on a scale from 0-1
- *Energy*: Measure from 0.0-1.0, perceptual measure of intensity and activity
- *Key*: The key the track is in using standard Pitch Class notation
- *Loudness*: overall loudness of a track in decibels
- *Mode*: indicates modality of a track (major or minor)
- *Speechiness*: detects presence of spoken words
- *Acousticness*: 0.0-1.0
- *Instrumentalness*: 0.0-1.0, predicts whether a track contains vocals
- *Liveness*: Detects the presence of an audience (0.0-1.0)
- *Valence*: 0.0-1.0 musical positiveness (happy vs. sad)
- *Tempo*: overall estimated tempo of a track in beats per minute
- *Duration_ms*: how long the track is in milliseconds
- *Time_signature*: an estimated overall time signature

What is interesting?

- It's interesting how this dataset includes many summarizing elements and traits of music that are not explicitly presented in Spotify's user interface. Although Spotify measures and determines these values constantly, they are not given or displayed. Some of the values are intrinsic to the music composition - such as tempo, key, mode, and time signature - while others are not arbitrary but more obscure measurement such as "danceability" or "liveness".

What is weird?

- Some of the measurements that are intuitively correlated with popular music (ex. danceability) have some low data points in the data set. The juxtaposition of all these qualities demonstrates how musical popularity is complex and multifaceted.