Music Consumption - Which Music Features Make Certain Genres More Popular Than Others Over Time?

Objective: Identify which genres and artists are most popular based on Spotify charts + Spotify metrics, which songs are most popular, what makes a song/genre popular… is speechiness/acoustiness/loudness/instrumentalness/valence/liveness/tempo/duration of song? I will analyze and deep dive into two datasets that span 1921-2023 to answer these interesting questions! How has music developed over time??? There's so much value to pry out in the data! I will leverage both datasets below, and I thoroughly outlined each dataset as well. Source 1 has over 1MM rows of data and 19 features while source 2 has almost 170k rows and 19 features. Each source has a lot of feature overlap too which is great.

Source 1: https://www.kaggle.com/datasets/amitanshjoshi/spotify-1million-tracks

-File = spotify_data.csv

About the Dataset:

This dataset was extracted from the Spotify platform using the Python library "Spotipy", which allows users to access music data provided via APIs. The dataset collected includes about 1 Million tracks with 19 features between 2000 and 2023. Also, there is a total of 61,445 unique artists and 82 genres in the data.

This clean data has been prepared and utilized for research purposes. Its significance lies in its potential to unravel patterns and predict song popularity prior to its release. This dataset could be used to create various predictive models with machine-learning/deep-learning techniques.

Audio Features	Description
Popularity	Track popularity (0 to 100)
Year	Year released (2000 to 2023)
Danceability	Track suitability for dancing (0.0 to 1.0)
Energy	The perceptual measure of intensity and activity (0.0 to 1.0)
Key	The key, the track is in (-1 to -11)
Loudness	Overall loudness of track in decibels (-60 to 0 dB)
Mode	Modality of the track (Major '1'/ Minor '0')
Speechiness	Presence of spoken words in the track
Acousticness	Confidence measure from 0 to 1 of whether the track is acoustic
Instrumentalness	Whether tracks contain vocals. (0.0 to 1.0)
Liveness	Presence of audience in the recording (0.0 – 1.0)
Valence	Musical positiveness (0.0 to 1.0)
Tempo	Tempo of the track in beats per minute (BPM)
Time_signature	Estimated time signature (3 to 7)
Duration_ms	Duration of track in milliseconds

File = spotify_data.csv

- 

- 

Source 2: https://www.kaggle.com/datasets/ektanegi/spotifydata-19212020

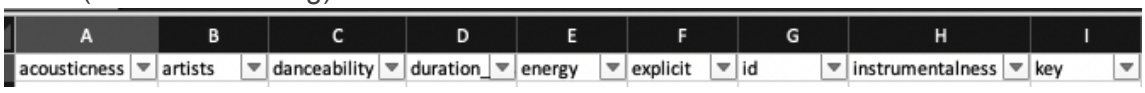
-File name = data.csv

About the Dataset:

The "data.csv" file contains more than 160.000 songs collected from Spotify Web API. The dataset is from Spotify and contains 169k songs from the year 1921 to year 2020. Each year got top 100 songs.

Primary:

- id (Id of track generated by Spotify)
  Numerical:
- acousticness (Ranges from 0 to 1)
- danceability (Ranges from 0 to 1)
- energy (Ranges from 0 to 1)
- duration_ms (Integer typically ranging from 200k to 300k)
- instrumentalness (Ranges from 0 to 1)
- valence (Ranges from 0 to 1)
- popularity (Ranges from 0 to 100)
- tempo (Float typically ranging from 50 to 150)
- liveness (Ranges from 0 to 1)
- loudness (Float typically ranging from -60 to 0)
- speechiness (Ranges from 0 to 1)
- year (Ranges from 1921 to 2020)
  Dummy:
- mode (0 = Minor, 1 = Major)
- explicit (0 = No explicit content, 1 = Explicit content)
  Categorical:
- key (All keys on octave encoded as values ranging from 0 to 11, starting on C as 0, C# as 1 and so on…)
- artists (List of artists mentioned)
- release_date (Date of release mostly in yyyy-mm-dd format, however precision of date may vary)
- name (Name of the song)



A	B	C	D	E	F	G	H	I
acousticness	artists	danceability	duration_	energy	explicit	id	instrumentalness	key

J	K	L	M	N	O	P	Q	R	S
liveness	loudness	mode	name	popularity	release_date	speechiness	tempo	valence	year