Trisha Karani
Databases
11/13/22

Project Description
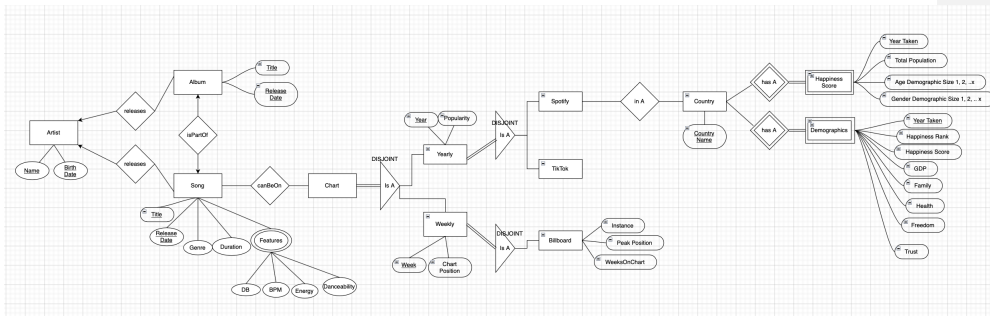
Domain

The domain we'd like to address in this project is analyzing music listening trends in different countries and its relationship to country demographics. We will be analyzing the top songs across various platforms (Billboard rankings, Spotify, Tiktok) and determining key characteristics of the songs and how they relate to the happiness, population, gender diversity in countries and using this to determine similarities between different countries.

Questions

- What is the most common genre across different countries?
- How many top TikTok songs in 2019 were top Billboard songs and top Spotify songs?
- What features of a song are common for Billboard songs?
    - Average energy?
    - Average danceability?
    - Average tempo?
    - Male or female singer?
- How has streaming changed the duration of songs?
    - What is the average duration of top Billboard songs?
    - What is the average duration of top Spotify songs?
    - What is the average duration of top Tiktok songs?
- Which artist appeared the most times on Billboard top songs?
- How does danceability of the top song in x country relate to happiness rank of the country?
- What are the demographics of countries where pop is the top genre?
    - Male to female ratio?
    - Largest age group?
- How did listening habits change during Covid-19 pandemic?
    - What was the most popular genre in 2019 vs. 2020 vs. 2021?
    - What was the average tempo of the top songs in 2019 vs. 2020 vs. 2021?
- In what years was hip hop genre on the charts?
- What percent of top charting songs included the word(s) "love"?
- Are Billboard charts representative of all popular songs?
    - What percent of top Spotify (streaming) songs are of a genre that is not on the Billboard charts?
    - Which songs were popular on Spotify (streaming) but not on Billboard?
    - How do the average features of the songs only popular on Spotify differ from the popular Billboard songs?
- How long do top TikTok songs last on the Billboard charts?

E-R Diagram

https://drive.google.com/file/d/130Ytsr3KlabEOwnhz3hbU_nFKstafw7x/view?usp=sharing

Questions
- How to put year in the ER diagram?
- Some datasets have a feature "popularity" (0-100) and others have "rank" (1-50)
  - Hard to combine these two
- Could possibly combine use datasets with weekly rankings to be daily rankings
-

Assumptions and Things to Note
- A Song can only be on one Album
- A Song can only be released by one Artist
- An Artist is considered to be an entity that releases music. Bands/groups are considered to be an Artist, and if individual members of the band release music, they are considered to be a separate Artist
  - Example: Harry Styles and One Direction will be separate Artists in the table
- All Songs are part of Albums

Limitations
- Datasets primarily focused on songs in English, so they may not be reflective of
- This project assumes that charts are representative of what songs/music types/genres are popular. However, the songs that are on charts are usually singles that were released by big record labels that have money and are also typically in the pop genre. One extension of this project may be to consider a way to evaluate songs that are popular but not on music charts.

Functional Dependencies
- releaseYear from Song < releaseYear from Album
  - Some songs are released as singles before the album it belongs on is released.

Commented [TK1]: Maybe get rid of this limitation? Some singles / popular songs may not be associated with an album.

-

Datasets

https://marianaossilva.github.io/DSW2019/assets/data/presentation.pdf

- Top Tiktok Songs
  - 2019
    - https://www.kaggle.com/datasets/sveta151/tiktok-popular-songs-2019
  - 2020
    - https://www.kaggle.com/datasets/sveta151/tiktok-popular-songs-2020
  - 2021
    - https://www.kaggle.com/datasets/sveta151/tiktok-popular-songs-2021
- Top 50 Spotify Songs
  - 2019
    - https://www.kaggle.com/datasets/leonardopena/top50spotify2019
  - 2010-2019
    - https://www.kaggle.com/datasets/leonardopena/top-spotify-songs-from-20102019-by-year
  - Daily worldwide ranking (2017-2018)
    - https://www.kaggle.com/datasets/edumucelli/spotifys-worldwide-daily-song-ranking
    -
- Billboard Hot 100 Songs
  - U.S.
    - https://www.kaggle.com/datasets/dhruvildave/billboard-the-hot-100-songs
    - https://github.com/HipsterVizNinja/random-data/tree/main/Music/hot-100
- Spotify song dataset
  - https://www.kaggle.com/datasets/mrmorj/dataset-of-songs-in-spotify
  - https://www.kaggle.com/datasets/rodolfofigueroa/spotify-12m-songs
- World Happiness Report
  - https://www.kaggle.com/datasets/unsdsn/world-happiness