Project Plan Proposal - HYP Songs

Main Goal: What affects media/song engagement?

Priority	Task	Dataset	Summary
0	Clustering	https://www.kaggle.com/code/nelgiriyewithana/an-introduction-to-top-spotify-songs-2023/input	Matching songs using use cases
1	Trendy Songs EDA/Analysis	https://www.kaggle.com/code/nelgiriyewithana/an-introduction-to-top-spotify-songs-2023/input	What makes a song trendy?
2	Sentiment Analysis	https://www.kaggle.com/datasets/saurabhshahane/music-dataset-1950-to-2019	Sentiment and song engagement
3	ML Feature Extraction	https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification/data	Using audio clips to extract features

Given we only have 8 weeks, we will focus on clustering, EDA of trendy songs, sentiment analysis and ML feature extraction. If time/resources permit, we will move on to a possible recommendation system and music generation for specific use cases as listed below.

Below are the methods we will use along with the goal we want to achieve by using them:

Clustering
Understanding the features that can be used to classify songs for different use cases (marketing campaign, background music, etc..)

Machine Learning:
We have proposed two methods in doing the machine learning model to extract spotify features, such as danceability and energy, and **we would like to ask which one would you prefer**.

1. Extract basic features (tempo, speed, loudness,etc.) from audio files using music libraries such as Librosa, then predict/derive spotify features (danceability, energy) based on the basic features that were extracted.

2. Train the model using the audio track for songs in spotify playlist and the spotify features (danceability, energy) as label. Then, predict spotify features directly from sound tracks for songs.

Sentiment Analysis :
Understanding the sentiment and theme of the song, i.e. positive, negative, neutral - what makes them positive, negative using either lyrics, musical/audio features or a combination of both

Trendy Song EDA: Exploration to get what makes a song trendy then if results yield insights move to predicting if a song will be trendy.

Music Generation
Generate songs based on use cases (party, marketing campaign).
https://huggingface.co/facebook/musicgen-stereo-melody-large

Recommendation system (maybe if we are able to) : Curate a playlist/suggest a song based on user specifications. Specifications could include musical attributes (eg. high energy song), use cases (marketing campaign, movie soundtrack,party) or sentiment (sad or happy song). We would need a very big dataset for this or we would have to make a dataset ourselves by pulling a lot of playlists from Spotify.
https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset?resource=download