DATA 602 Final Project Proposal
Author: Kevin Kirby

Research Question:
How can retrieval-augmented semantic search enhance song recommendation systems by incorporating song metadata, artist biographies, genre characteristics, and song popularity?

Context:
In spring and summer, I used advanced machine learning techniques to evaluate a Kaggle Spotify and Beatport dataset, building a small neural network with TensorFlow (details here). This proposal builds on that work, reimagining how I use the data from the start.

Justification: I work as a data product manager in Disney's ad tech division, building data products to enhance ads on Disney Plus and Hulu, improving both consumer experiences and advertiser outcomes. Due to conflict of interest rules, I can't work on video streaming projects, but I want to explore advanced deep learning with music data, which closely parallels movie and TV data

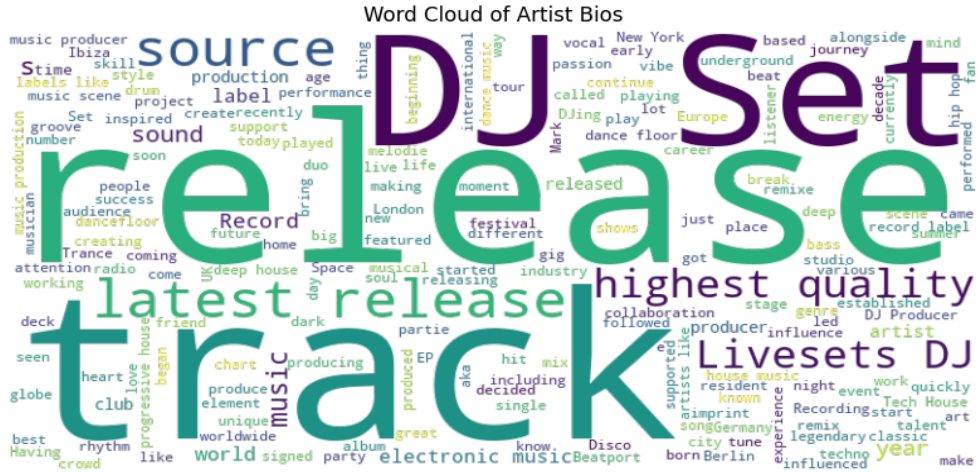
Data Sources
The Kaggle dataset**;** Beatport.com artist bios; Beatstats points. Scraping Fish web scraping

Required libraries:bs4; pandas, numpy, tensorflow; transformers; langchain; sentence-transformers; wandb;

Meta's Llama 3.3 11B parameter model: I have a working connection to the model via Hugging Face. I intend to use it for this project

EDA:
I tried to keep this whole thing to one page but there's some spillover. The spillover is just charts though.

Word cloud using WordCloud library. This is based on 55,000 artist bios I scraped fro Beatport:



All of the below charts are based on a combination of song metadata and artist bios. I wanted to understand the worlds being used and their context.