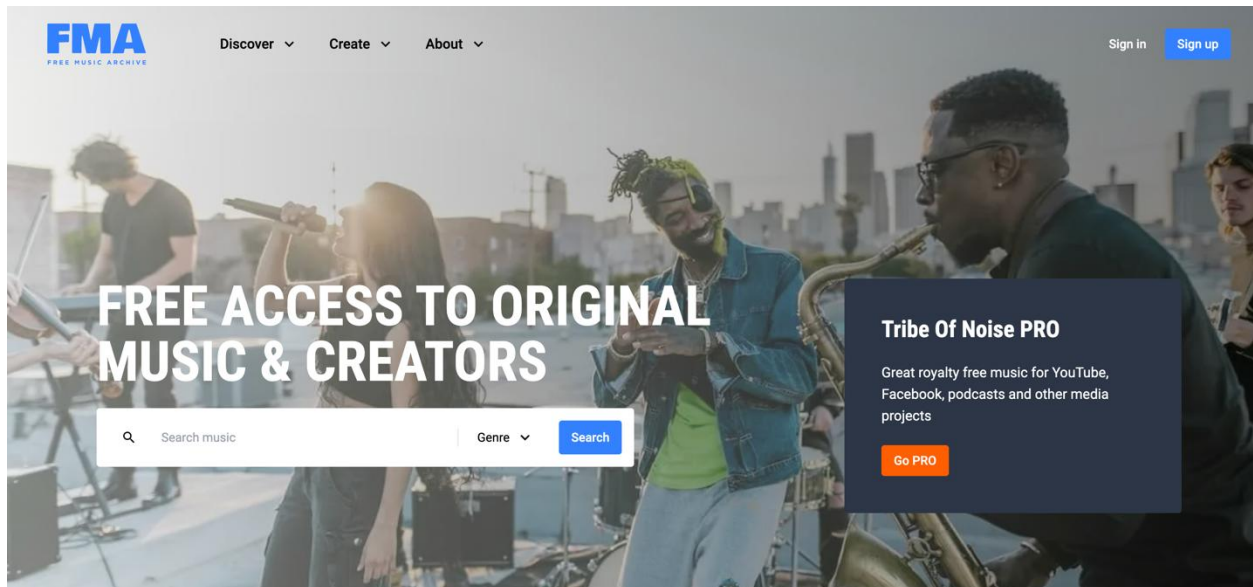# HW Assignment 4 (A4) Part 1: Description of FMA data
CS6140: Machine Learning Fall 2024
Due Date: Thursday, _____
(5 points)

Objective

This assignment is designed to familiarize you with the Free Music Archive (FMA) dataset, focusing on two main data files: **features.csv** and **tracks.csv**. You will gain a deeper understanding of the structure and content of these files, which is essential for effective data exploration and analysis.

1) Dataset Overview

Begin with a brief description of the FMA dataset. Discuss its purpose and the type of data it contains, such as audio features and metadata for a diverse collection of music tracks.

2) Describing Features in features.csv

Domain Analysis: The features.csv file consists of various domains of audio features, such as chroma and MFCC. There are **9 domains** in total. For each domain:

- Provide a description of what the domain represents in the context of audio data (e.g., chroma, MFCC).

- Include the total number of features within each domain. For example, *chroma* contains 252 features

3) Describing Metadata in tracks.csv

- Choose a few key columns from the metadata section in tracks.csv, such as track title, artist, album, and genre and others.
- Provide a brief description of each chosen column, detailing the type of information it holds.

Data Integration Plan: Describe a strategy for efficiently joining features.csv and tracks.csv to facilitate future analysis. Consider factors like unique identifiers and any relevant columns that can optimize the join process.

Genre Frequency Distribution

Using tracks.csv, identify and describe the frequency distribution of the available genres. You can present the frequency distribution in a table or chart format if desired. This will provide insights into the diversity of genres present in the dataset.