Music Genre Classifier

Applied Machine Learning - Project Deliverable #2

Group 22 - James Potash, Xena Maayah, Rohan Poddar, Yijun Zhao, Noame de Boerdere

Project Overview

With the ever-increasing digitalizing of music from creation to distribution and streaming platforms, the amount of music available has exploded in recent years. This has made it more challenging to manage and navigate through the vast collection of music available.

To create a better and easier user experience, companies like Spotify, Apple Music, Soundcloud, and others have utilized machine learning to generate recommendation and search systems to provide users with a more personalized and relevant experience based on their preferences. These companies use music genre classifiers as part of their recommendation system.

Our group will create a music genre classification system in this project.

Dataset Used

GTZAN Music Classification Dataset - A collection of around 10000 30-second-long audio files split into 10 genres. The genres are - blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock.

Audio Files

- 1000 30-second long audio files split into ten folders, 1 for each genre. The audio files are of type .wav

Mel Spectrograms

- The audio files in the dataset have been visualized using Mel Spectrograms, i.e., a visual representation of the frequency content of a signal that has been transformed using a Mel scale.

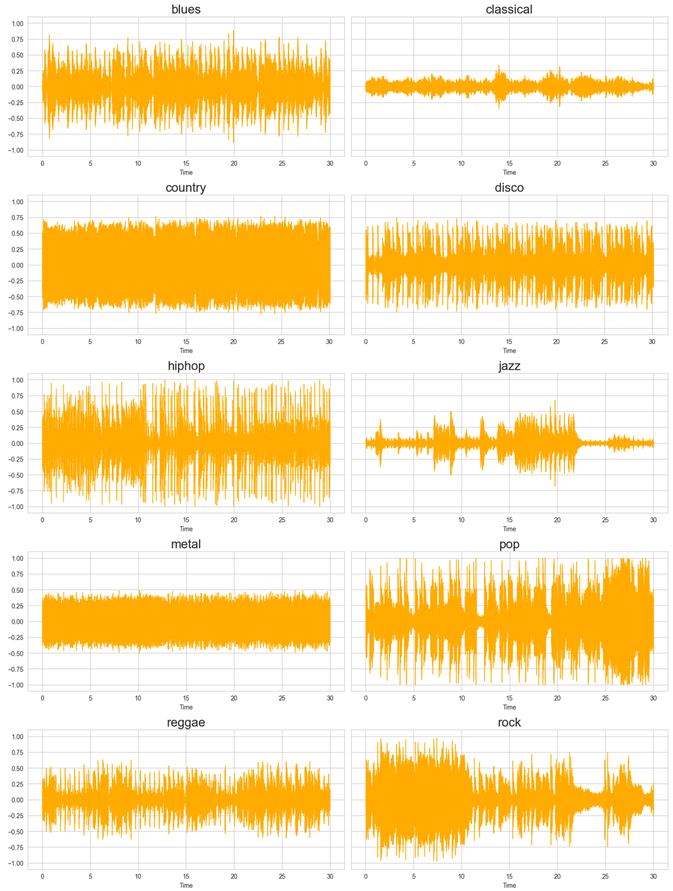
Features - 3 secs

- Similar to the Features – 30-sec file, however, the audio files are divided into 3-second segments, and then the features are calculated. This gives us more data points to use.

Initial Data Exploration and Visualization

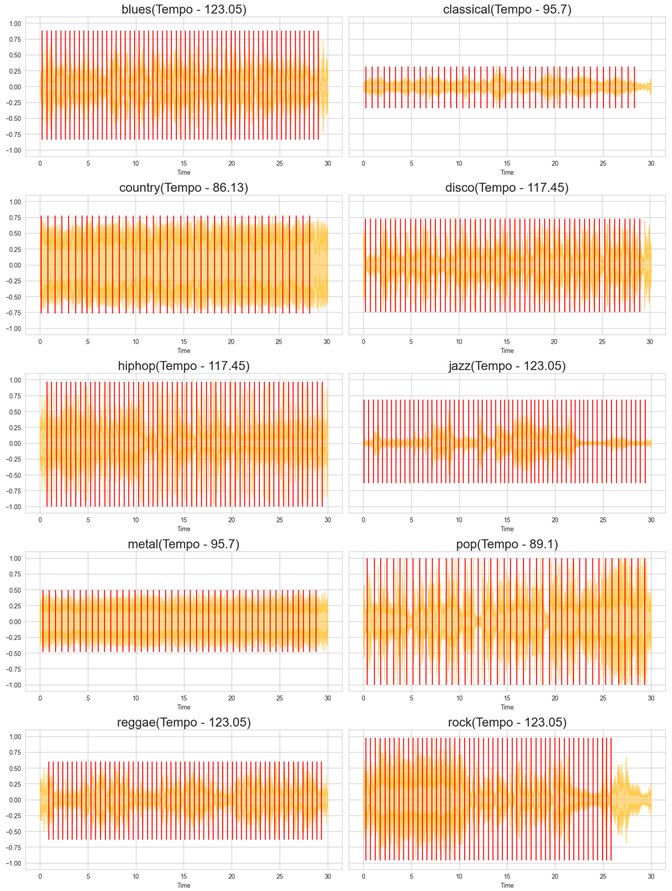
Audio Files

Wave form



Wave form + Spectral Centroid

Wave from + Tempo

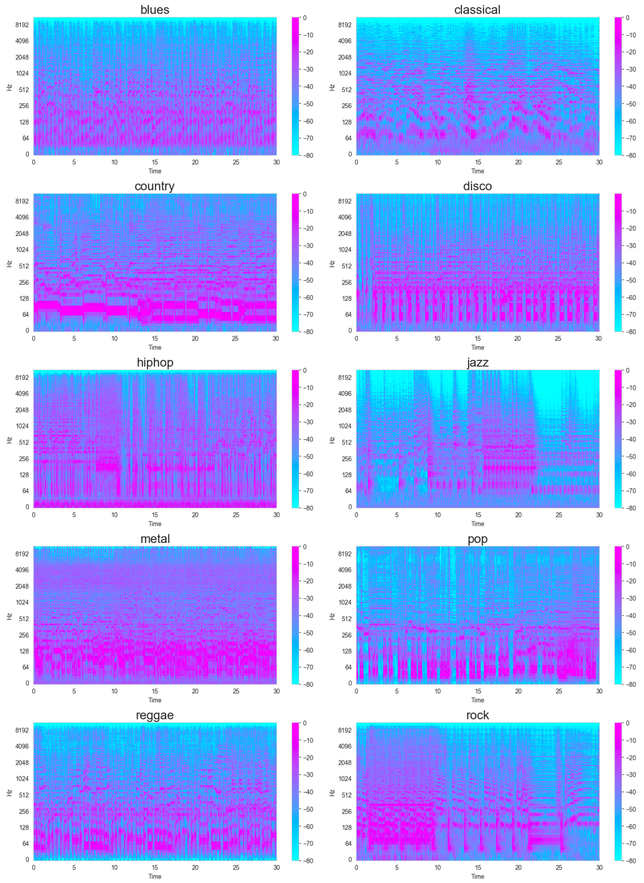


Visualization of the first track of each genre

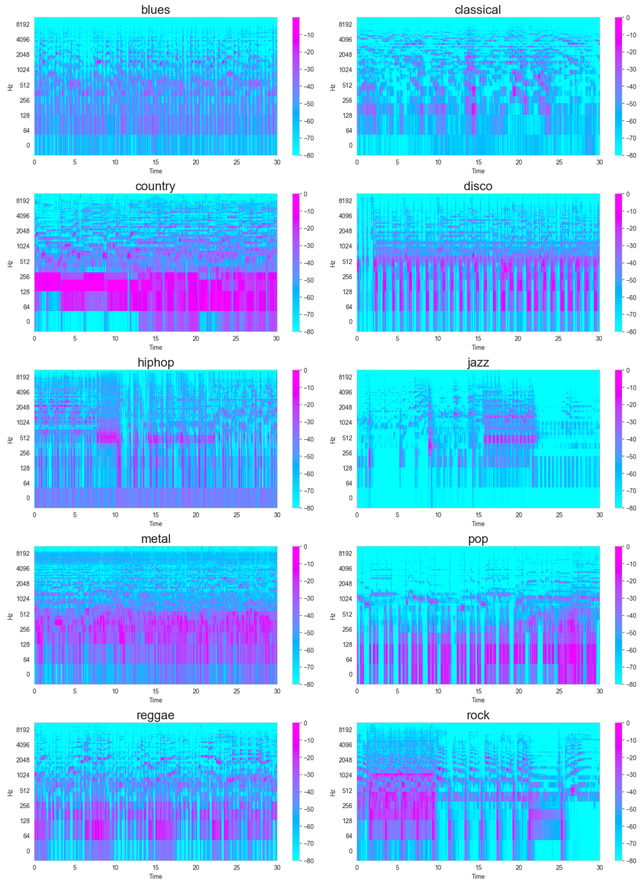
Initial Data Exploration and Visualization

Mel Spectrogram

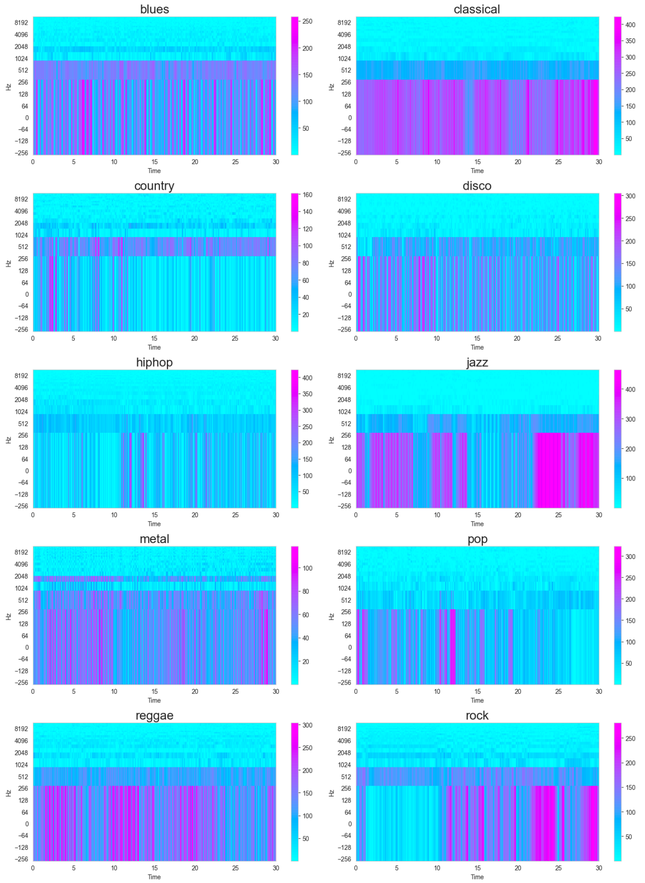
Amplitude Spectrogram



Mel Spectrogram



Mel-frequency cepstral coefficients (MFCCs)



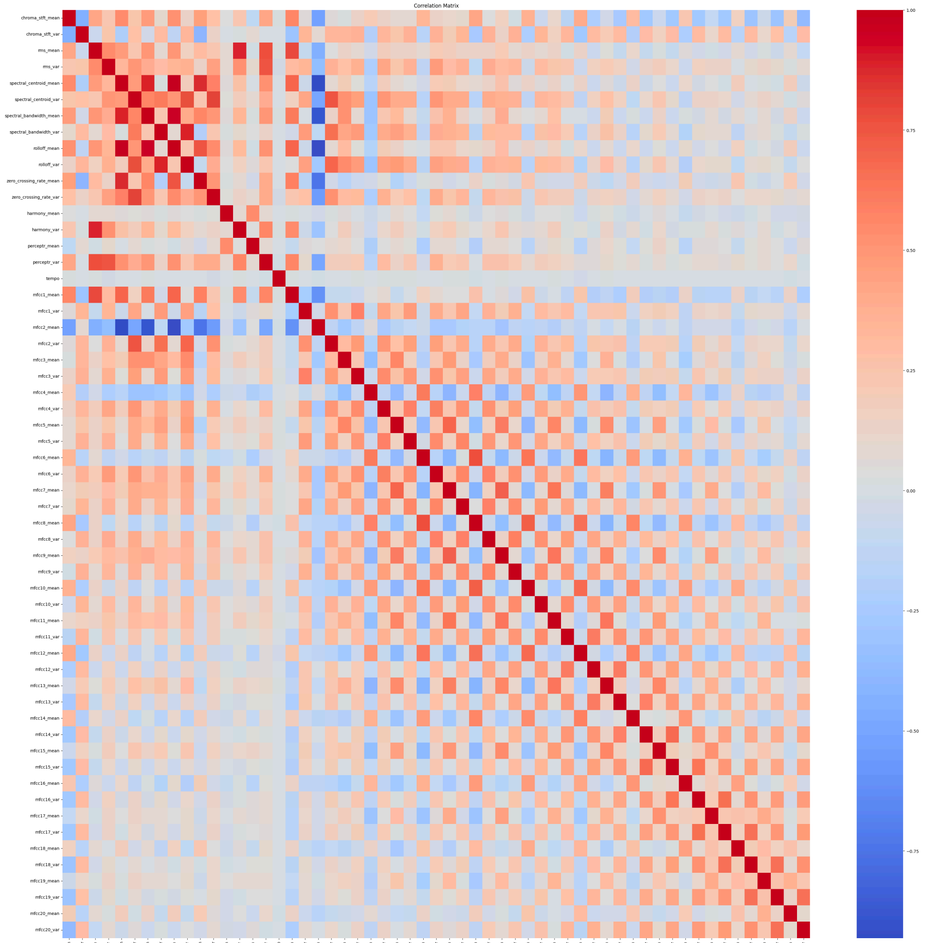
Visualization of the first track of each genre

Initial Data Exploration and Visualization

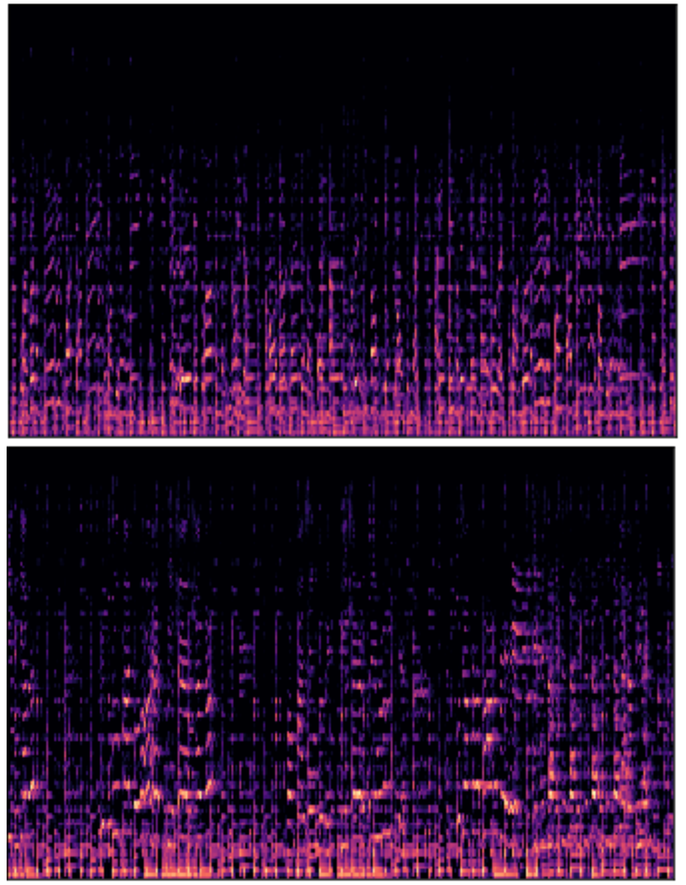
Features - 3 secs

Column Name	Type	Nulls
length	int64	0
chroma_stft_mean	float64	0
chroma_stft_var	float64	0
rms_mean	float64	0
rms_var	float64	0
spectral_centroid_mean	float64	0
spectral_centroid_var	float64	0
spectral_bandwidth_mean	float64	0
spectral_bandwidth_var	float64	0
rolloff_mean	float64	0
rolloff_var	float64	0
zero_crossing_rate_mean	float64	0
zero_crossing_rate_var	float64	0
harmony_mean	float64	0
harmony_var	float64	0
perceptr_mean	float64	0
perceptr_var	float64	0
tempo	float64	0
mfcc1_mean	float64	0
mfcc1_var	float64	0
mfcc2_mean	float64	0
mfcc2_var	float64	0
mfcc3_mean	float64	0
mfcc3_var	float64	0
mfcc4_mean	float64	0
...		
mfcc20_mean	float64	0
mfcc20_var	float64	0
genre	object	0

Correlation Matrix of Numerical Columns



Mel Spectrograms



3 color channels - RGB
Image Size - 432 × 288 pixels

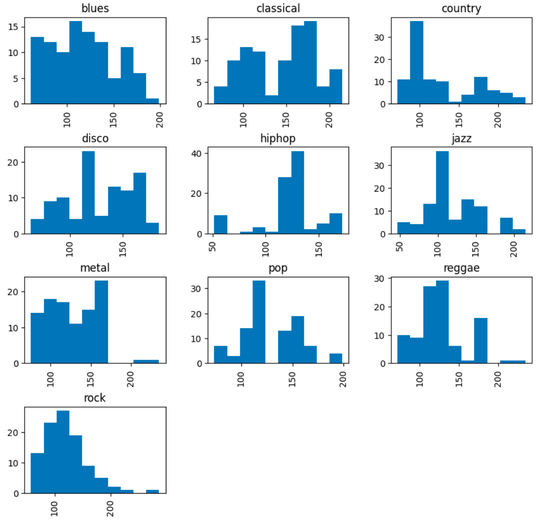
Initial Data Exploration and Visualization

Features - 3 secs

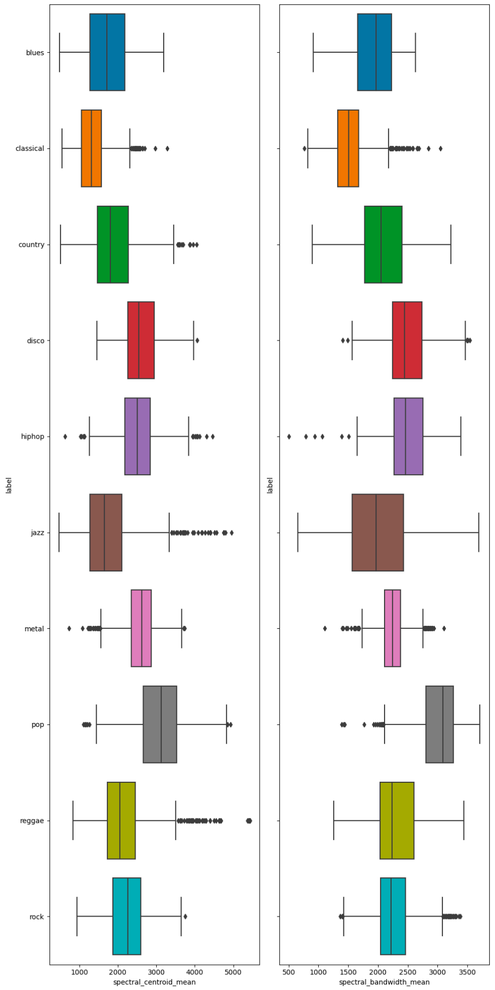
Number of datapoints by Genre

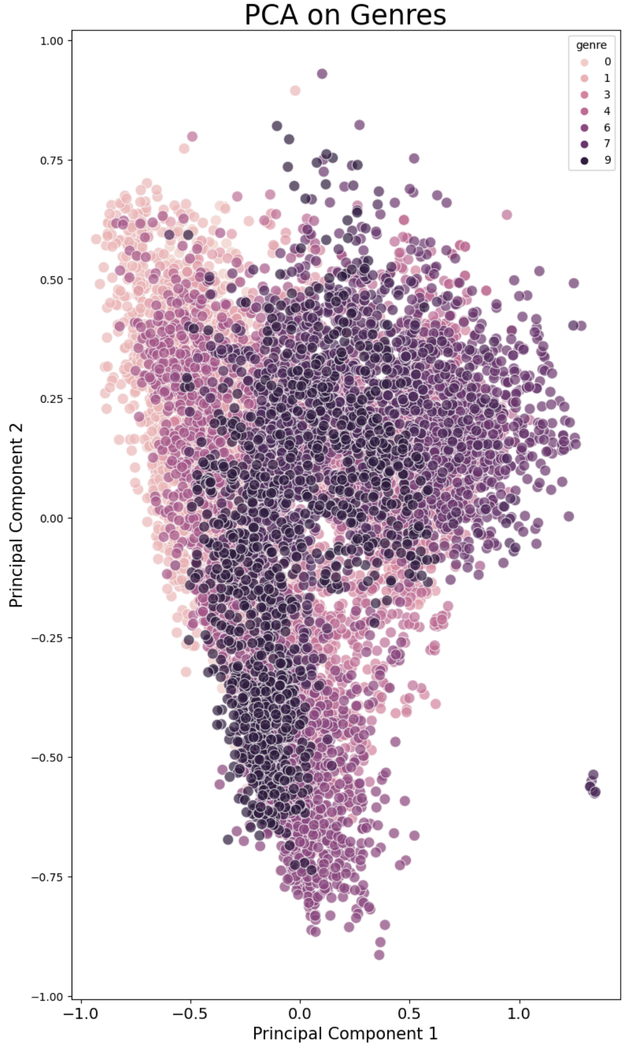
	Features - 3s	Audio Files
Genre	**# Data**	**# Data**
classical	998	100
country	997	100
disco	999	100
hiphop	998	100
jazz	1000	100
metal	1000	100
pop	1000	100
reggae	1000	100
rock	998	100
blues	1000	100

Distribution of Tempo by Genre



Spectral Centroid by Genre





Cleaning and Sampling

- The dataset has no null values, so no further handling is required
- The data points are evenly distributed among the ten genres, so we have a balanced dataset. We will experiment with both random and stratified sampling and compare model performance.
- A few columns are highly correlated and have been dropped before training
- All the numerical features have been scaled and standardized
- Since the target variable is of the type object, we have mapped it to integer values for modeling.
- The image files have been resized, and the pixel values normalized before passing through the model.

Insights

Features - 30 secs

- We have a balanced dataset with no null values.
- There are some observable trends in the numerical features when we plot them by genre. This means the different classes have specific characteristics that can be learned well.
- The PCA in 2 dimensions doesn’t give us a good idea, but higher dimensions show some amount of separation between the points.
- Spectral Centroid, Spectral Bandwidth, and Roll off are highly correlated, which makes sense since these features explain similar characteristics of the audio wave. Similarly, harmony and rms are also highly correlated

Audio and Mel Spectrograms

- There are several ways to visualize the different components of an audio file, which could help the model learn various aspects.
- Visualizing the waveforms, spectral centroids, tempos, etc., shows a difference among the genres. A similar trend can be seen in the Mel Spectrograms as well.
- Mel Spectrograms capture more information than visuals of waveforms or waveforms superimposed with other audio wave characteristics. However, does this translate into better predictions if we use Mel Spectrograms over waveforms? This is something that we will explore during the modeling process.

Proposed Machine Learning Techniques

- For the numerical features of the audio files that are available in Features - 3 secs, we will use several Machine Learning techniques such as:

    1. Support Vector Machine (SVM)
    2. Random Forrest Classifier
    3. XGBoost

- For the Mel Spectrogram and other such image files, we will use a Convolutional Neural Network (CNN) for classification. We will compare the performance between a model trained from scratch on the data and a model that uses transfer learning on a pre-trained CNN such as ResNet18 or InceptionV3.