Music Genre Classification

Wen Zhang, Xinyi Zhang, Rujue Du

Table of contents

01

Introduction

- Business Problem
- Project Goal

Business Problem

Music has been known to humans for times immemorial. Ever since the melody of instruments has fallen on the ears of humans, it has implanted the seeds of emotions that are otherwise hard to achieve.

While music is altogether a melody that connects all hearts in this world, there are further classifications of music that bind together music lovers and keep the melody going on.

From K-Pop to Jazz, music lovers rely on the technology of music genre classification and are able to listen to songs as per their preferences. While it takes only a click for a listener to switch from Jazz music to Rap, there is certainly much more beneath the surface that fuels our love for music.

Goal:

- Build machine learning and deep learning models to classify music into different genres

Assumptions:

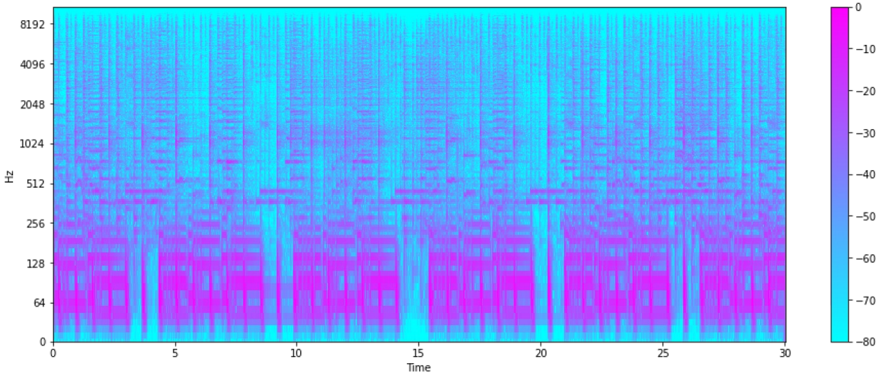
- Each data represents only one type of song

02

Data Analysis & Visualization

- Data Profile & Analyze
- Exploratory Data Analysis (EDA)

Data Profile (audio file)



Overview

Music genre data contains visual representation for each audio file collected from sources including personal CDs, radio, microphone recordings, etc.

Dimension

Collection with 10 genres (blues, classical, country, disco, hiphop, jazz, mental, pop, reggae, rock) with 100 audio files for each genre

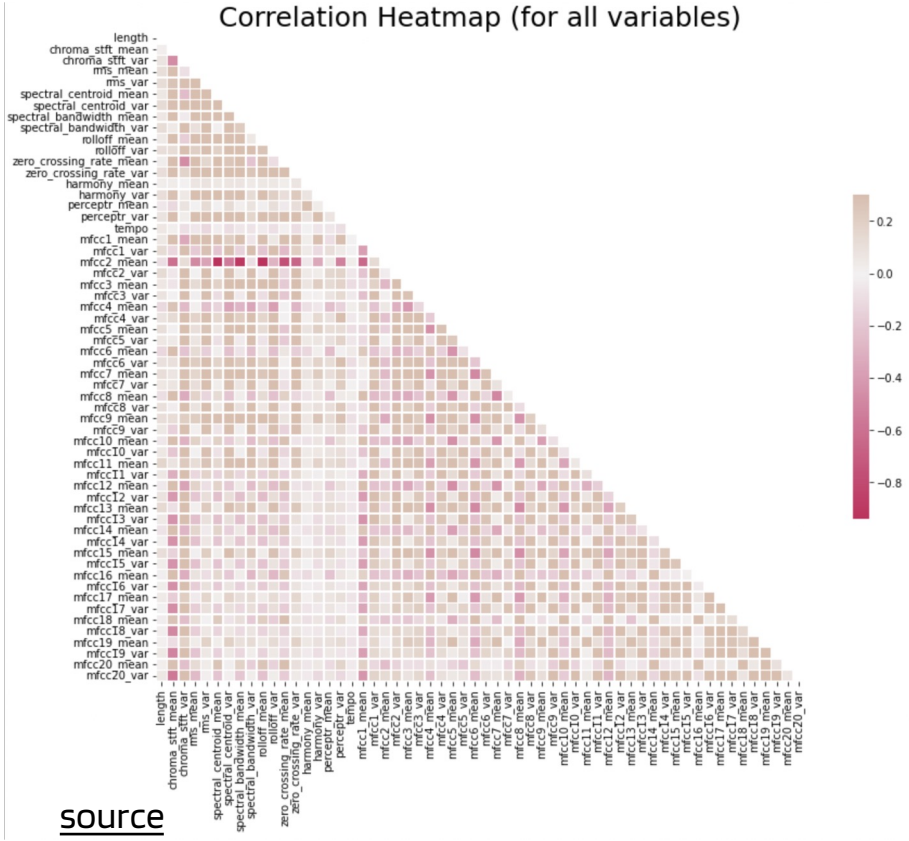
Formart

Melspectrograms for each audio saved as .wav file

Length & Size

30s audio each, with the image shape of (300, 300, 3)

Data Profile (csv file)



source

Overview

Music genre data with extracted features from audio file

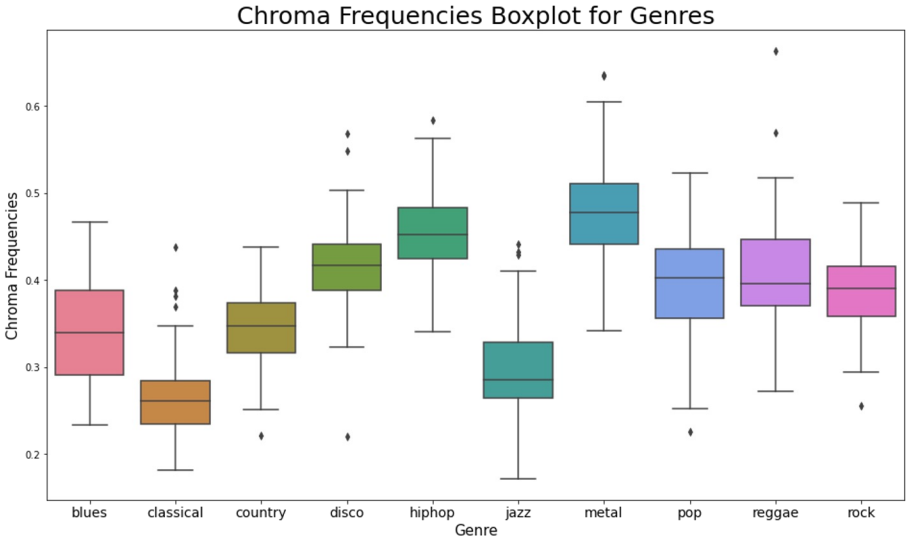
Dimension

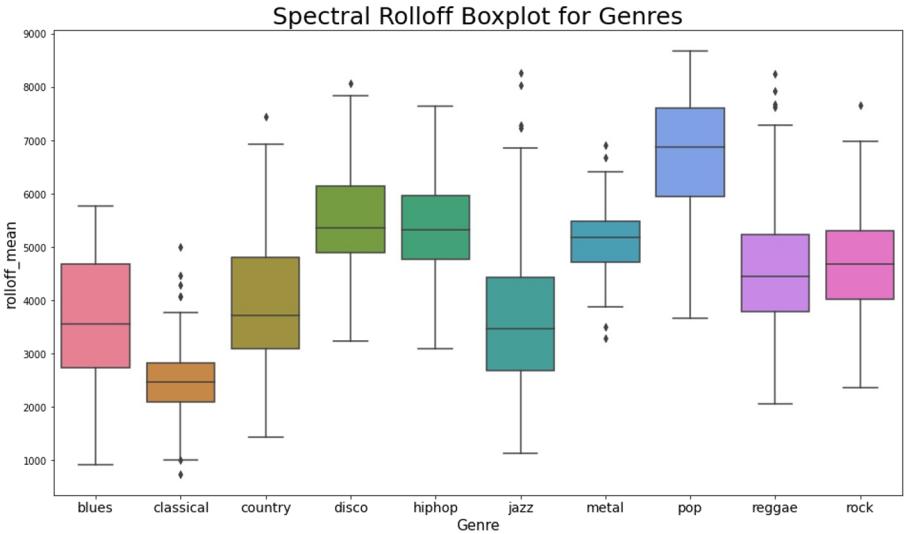
1000 audio samples with 60 features

Data Quality

No missing data, all fields are checked and formatted appropriately

EDA





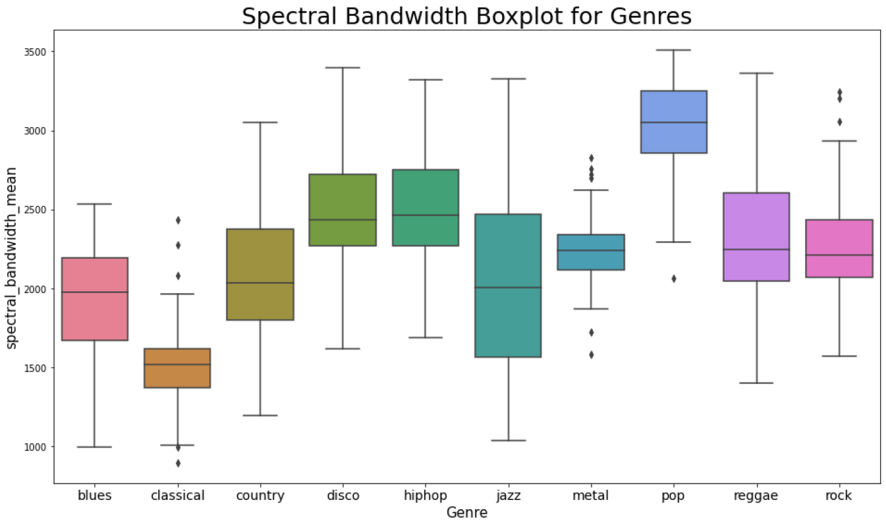
Chroma Frequencie (mean) vs Genre

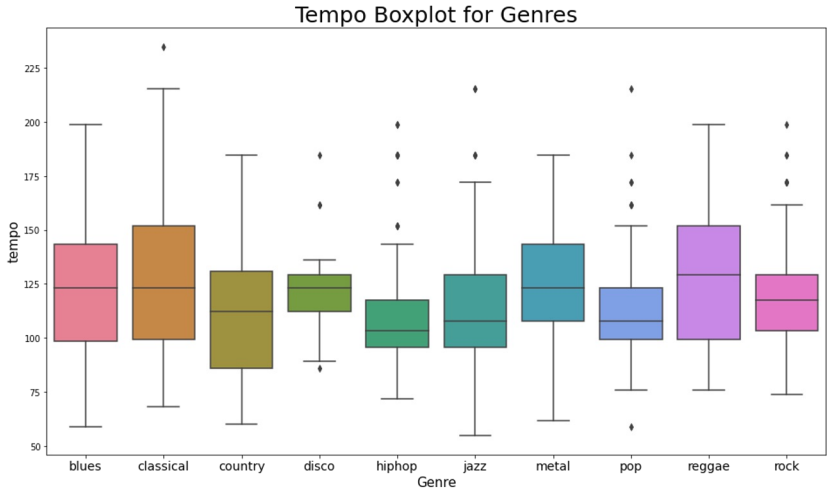
- CF is highly correlated with types of genre
- Classical and jazz tend tend to have low CF value

Spectral Rollloff (mean) vs Genre

- SR is highly correlated with types of genre
- Pop music tend to have high SR value

EDA





Spectral Bandwidth (mean) vs Genre

- SB is highly correlated with types of genre
- Pop music tend to have high SB value while classical music tend to have low SB value

Tempo vs Genre

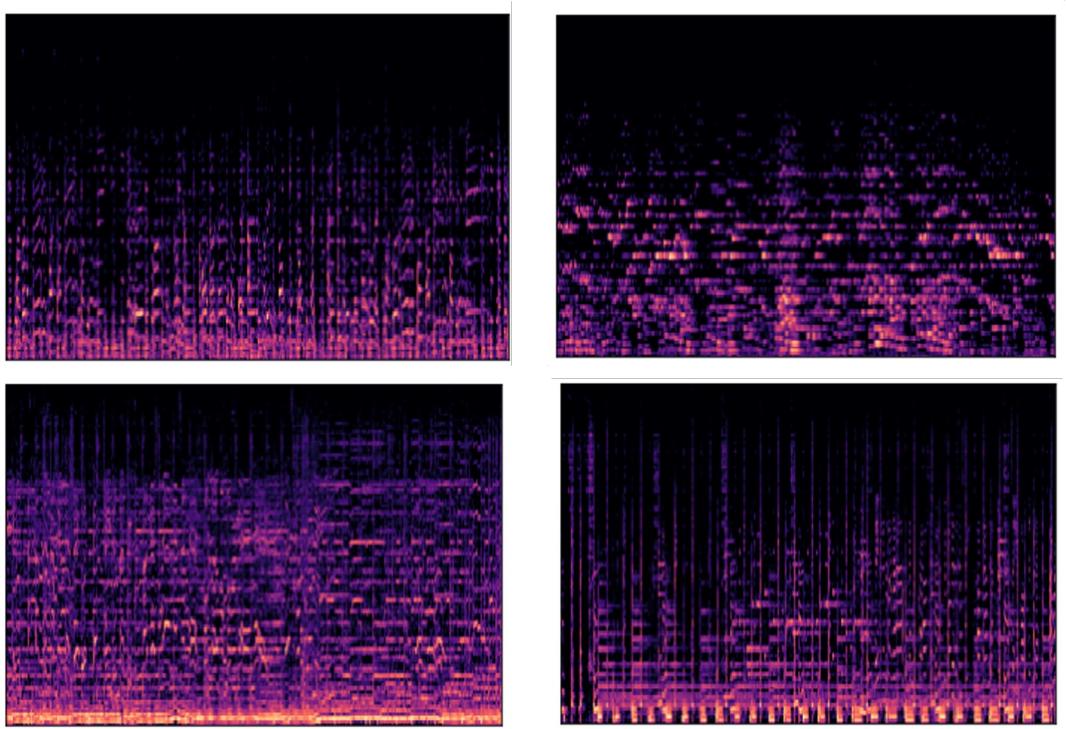
- Tempo, as one of the most common features of music, surprisingly, is not an obvious indicator for the type of music
- Most types have same median of tempo with only a few extreme cases

03

Model Evaluation

- CNN on Image
- CNN on Data
- ML Models on Data

Image Data



Top two are Blues, Classical, buttom two are Country, Disco
Hard to tell the difference through eyeballing

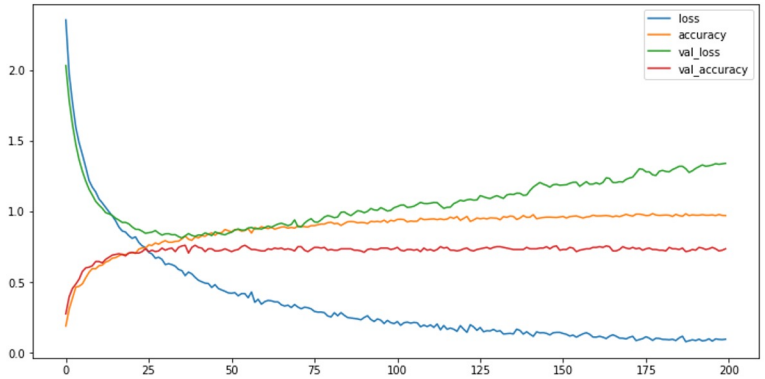
- **Train-Test Split**
  - Split Ratio: 80:20
  - 800 train vs. 200 test
- **Model Used:**
  - CNN (Convolutional Neural Network)
  - 2D Convolutional layer
    + 2D MaxPooling layer
    + Flatten layer
    + Dense layer

CNN ON Image

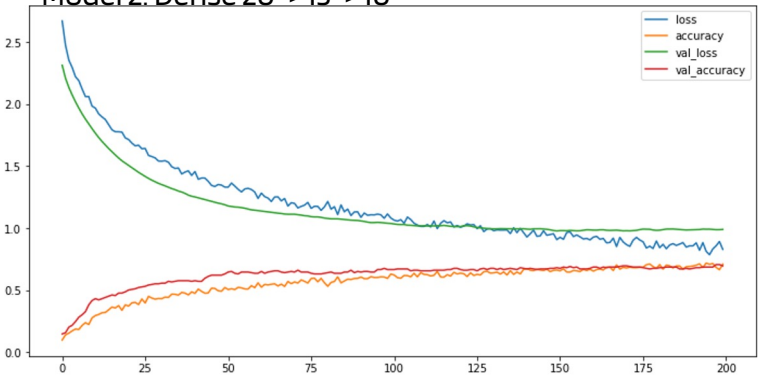
	Model Summary	Accuracy (Train)	Accuracy (Test)	Model Loss (Train)
Model 1	VGG16 : 5 blocks - 13 Conv2D + 5 MaxPooling2D +1 Flatten + 3 Dense	92.99%	24.5%	0.2735
Model 2	2 Conv2D (64 & 32) + 2 MaxPooling2D + 1 Flatten + 2 Dense	99.75%	43%	0.0132
Model 3	2 Conv2D (64 & 32)+ 2 MaxPooling2D + 1 Flatten + 1 Dense	99.87%	44%	0.0076
Model 4	1 Conv2D (64)+ 1 MaxPooling2D + 1 Flatten + 1 Dense	99.75%	35%	0.0100
Model 5	2 Conv2D (64 & 128)+ 2 MaxPooling2D + 1 Flatten + 1 Dense	99.87%	47.5%	0.0078

CNN on Data

Model 1: Dense 128->64->10



Model 2: Dense 20->15->10



	Training Loss	Testing Loss	Training Accuracy	Testing Accuracy
Model 1	0.0153	1.3393	99.87%	73.5%
Model 2	0.5201	0.9891	84.62%	69.5%

ML Models ON Data

Model Select	Accuracy (Train)	Accuracy (Test)
Naive Bayes	48.38%	43%
Stochastic Gradient Descent	79.63%	69%
KNN	74.63%	65.5%
Random Forest	99.88%	74%
Support Vector Machine	88.75%	71.5%
XG Booster	99.88%	74.5%

- SVM Model outperformes other models
- Most models contain problem of overfitting(Seriouse Overtting model: RF, XG Boost)
- KNN Model also has relatively good performance

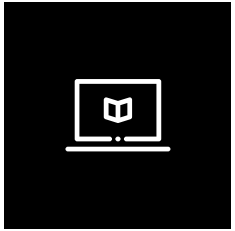
04

Conclusion & Future Work

Conclusions

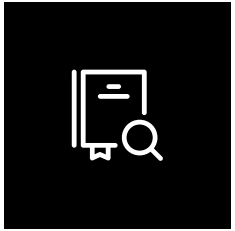
- The best model for all three parts (CNN on image, CNN on data, ML model on data) is SVM model, although the model does not reach the highest accuracy on both train and test dataset; However, it achieves the best balance on overfitting and model accuracy
- CNN for image classification has low accuracy on test. Due to the limited information on image compare to csv file, its overfitting issue is more serious
- Machine learning models performed better in avoiding overfitting issue, compare to CNN models

Future Work



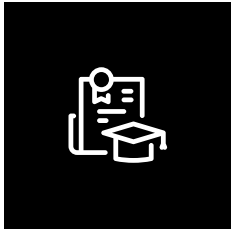
Fix Overfitting

- Add more data
- Use data augmentation
- Use architectures that generalize well
- Add regularization
- Reduce architecture complexity



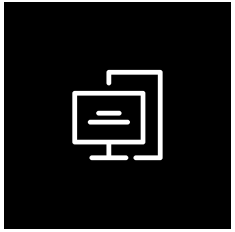
Audio Info Transformation

Consider to use other wave images or combine them with the original images(chroma frequency gram, spectrogram, sound waves) for NN models to see if there is a better performance



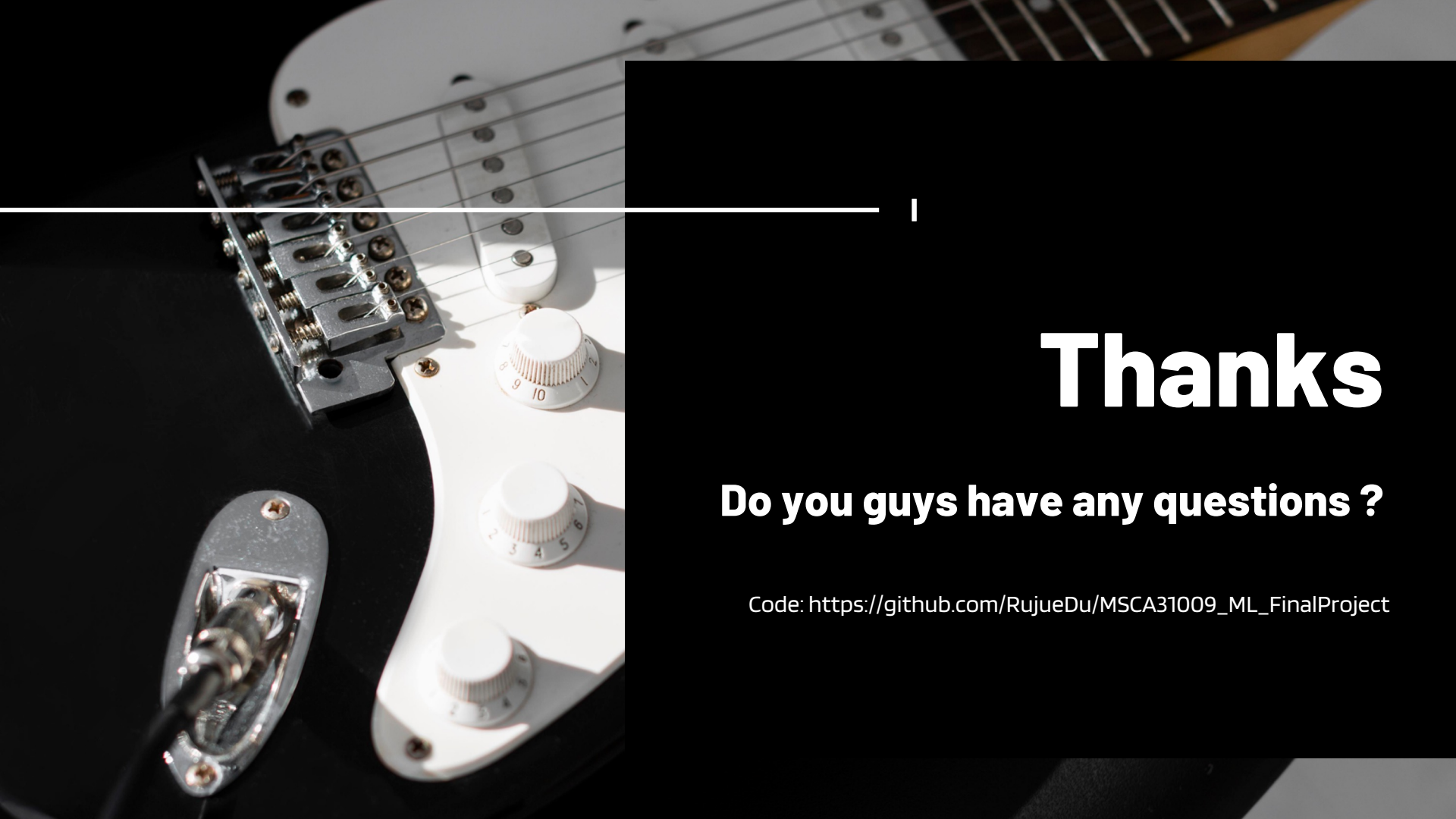
Model Combination

Combine CNN models with both csv and image data to capture more information contains in audios



New ways in fitting Image Model

Try to rotate, flip or stretch images to include more information on image classification



Thanks

Do you guys have any questions ?

Code: https://github.com/RujueDu/MSCA31009_ML_FinalProject

Appendix

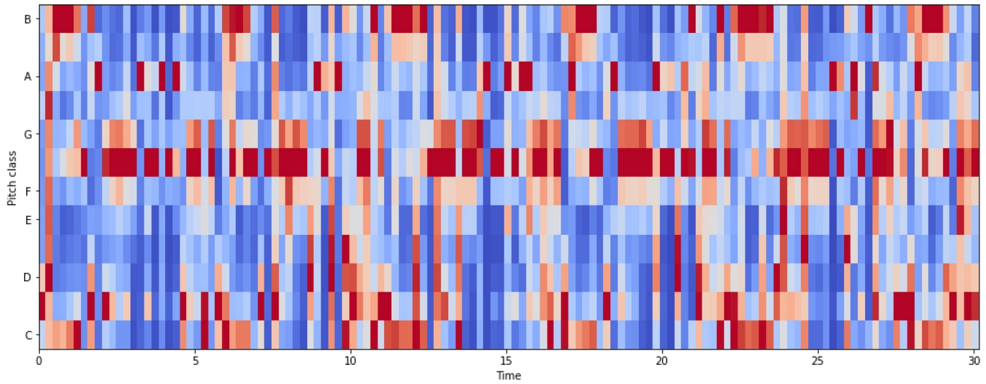
- Attribute Information
- Audio 2D visualization

Attribute Information

Chroma frequency

Chroma features are an interesting and powerful representation for music audio in which the entire spectrum is projected onto 12 bins representing the 12 distinct semitones (or chroma) of the musical octave.

Chromogram
(Sample Audio Randomly Selected: reggae.00036.wav)



Attribute Information

Spectral Centroid

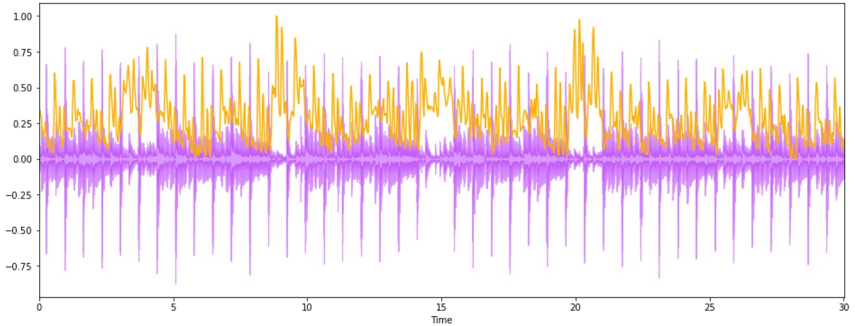
- Spectral Centroid indicates where the "centre of mass" for a sound is located and is calculated as the weighted mean of the frequencies present in the sound.

Spectral Bandwidth

- The Wavelength interval in which a radiated spectral quantity is not less than half its maximum value. It is a measure of the extent of the Spectrum For a Light Source typical spectral widths are 20 to 60 nm for a LED and 2 to 5 nm for a Laser diode.

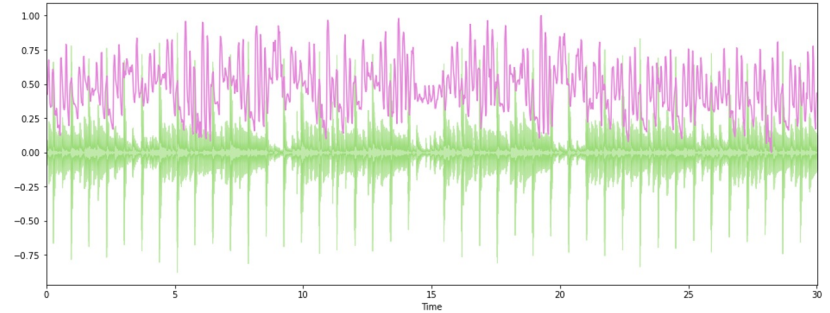
Spectral Centroid along the waveform

(Sample Audio Randomly Selected: reggae.00036.wav)



Spectral Bandwidth along the waveform

(Sample Audio Randomly Selected: reggae.00036.wav)



Attribute Information

Spectral Rolloff

- Spectral rolloff is the frequency below which a specified percentage of the total spectral energy

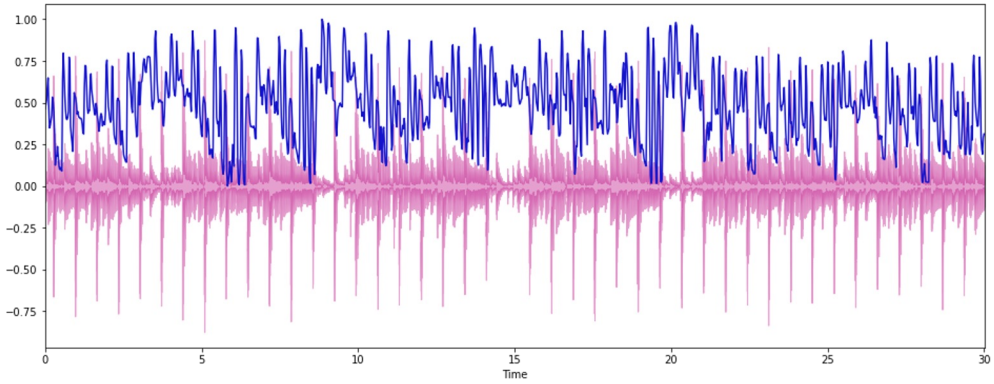
Zero Crossing Rate

- The rate at which the signal changes from positive to negative or back.

Tempo BMP (beats per minute)

- Tempo is a dynamic programming beat tracker

Spectral Rollover along the waveform
(Sample Audio Randomly Selected: reggae.00036.wav)



Attribute Information

Harmonics

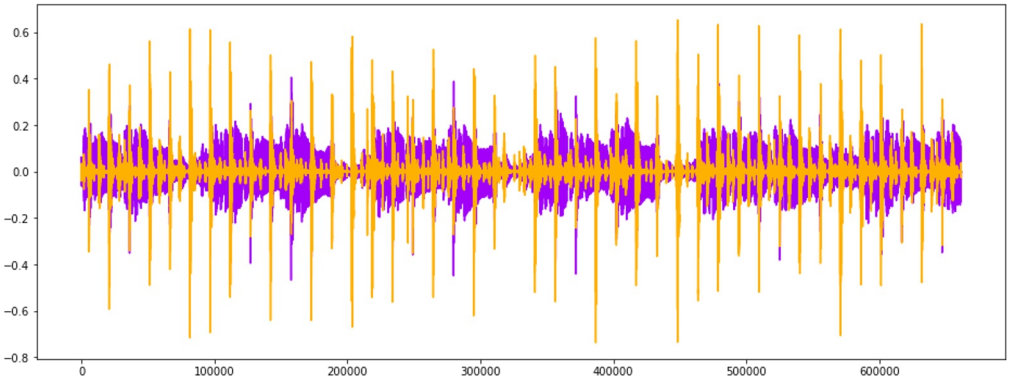
- Harmonics are characteristichs that human ears can't distinguish (represents the sound color)

Perceptrual

- Perceptrual understands shock wave, represents the sound rhythm and emotion

Harmonics and Perceptrual Plot

(Sample Audio Randomly Selected: reggae.00036.wav)



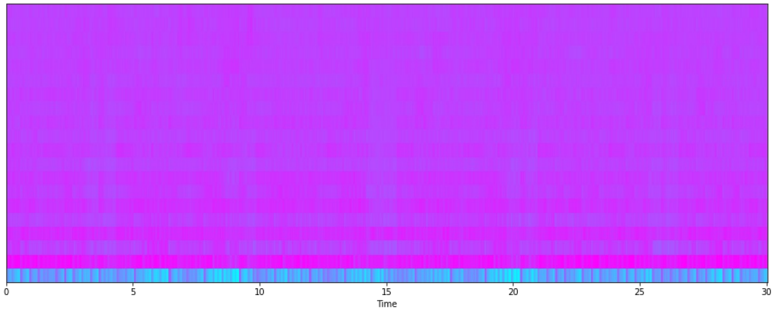
Attribute Information

Mel-Frequency Cepstral Coefficients(MFCC)

- In sound processing, the mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC. In the csv file, the extracted feature contains MFCC from order 1 to 20.

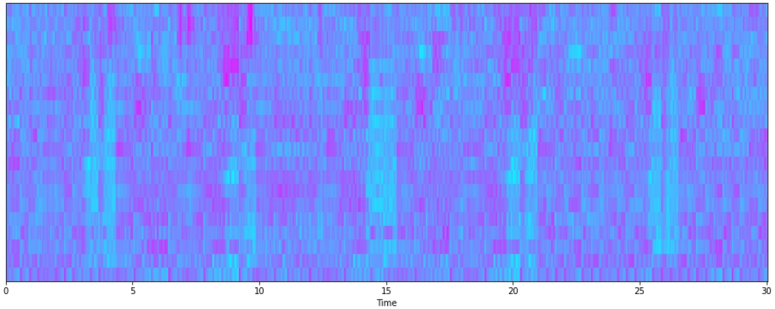
Original MFCC Plot

(Sample Audio Randomly Selected: reggae.00036.wav)



Scaled MFCC Plot

(Sample Audio Randomly Selected: reggae.00036.wav)

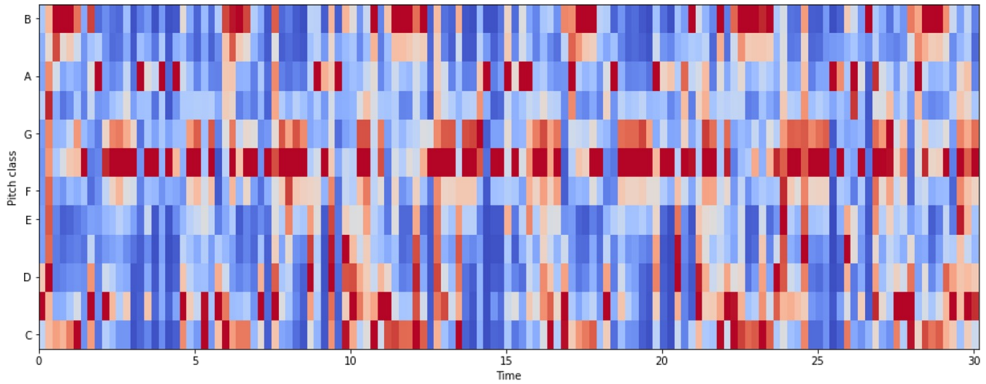


Attribute Information

Chroma frequency

Chroma features are an interesting and powerful representation for music audio in which the entire spectrum is projected onto 12 bins representing the 12 distinct semitones (or chroma) of the musical octave.

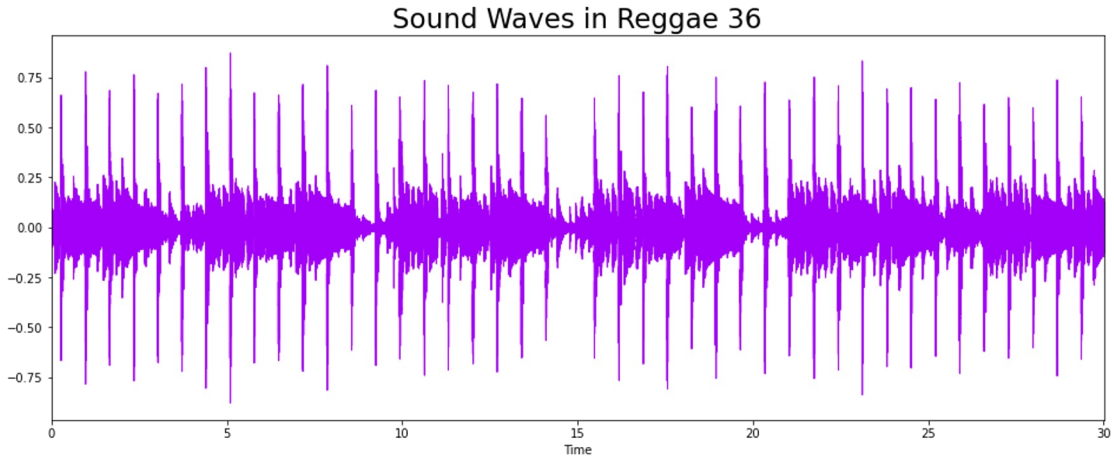
Chromogram
(Sample Audio Randomly Selected: reggae.00036.wav)



01

Sound Waves

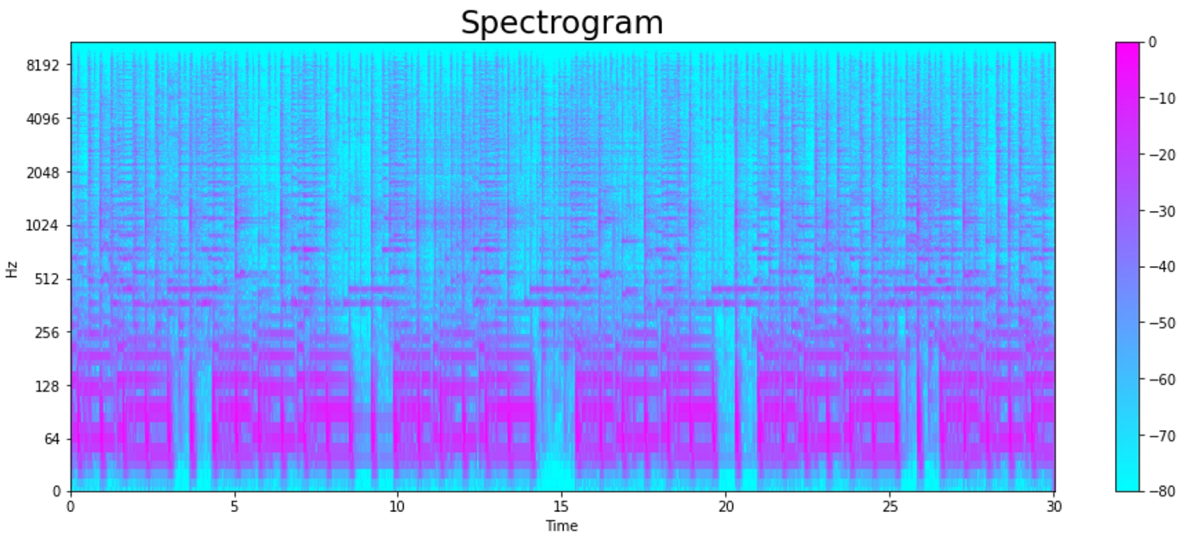
Sample Audio Randomly
Selected: reggae.00036.wav



02

Spectrogram

A visual representation of the spectrum of frequencies of a signal as it varies with time. Also called sonographs, voiceprints, or voicegrams when applied to an audio signal.



03
Mel

Spectrogram

The Mel Scale, mathematically speaking, is the result of some non-linear transformation of the frequency scale. The Mel Spectrogram is a normal Spectrogram, but with a Mel Scale on the y axis.