

```
In [1]: pip install pandas
```

```
Requirement already satisfied: pandas in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (2.2.2)  
Requirement already satisfied: numpy>=1.23.2 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from pandas) (1.26.4)  
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from pandas) (2.9.0.post0)  
Requirement already satisfied: pytz>=2020.1 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from pandas) (2024.1)  
Requirement already satisfied: tzdata>=2022.7 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from pandas) (2024.1)  
Requirement already satisfied: six>=1.5 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)  
Note: you may need to restart the kernel to use updated packages.
```

```
In [2]: pip install numpy
```

```
Requirement already satisfied: numpy in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (1.26.4)  
Note: you may need to restart the kernel to use updated packages.
```

```
In [1]: pip install seaborn
```

Requirement already satisfied: seaborn in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (0.13.2)

Requirement already satisfied: numpy!=1.24.0,>=1.20 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from seaborn) (1.26.4)

Requirement already satisfied: pandas>=1.2 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from seaborn) (2.2.2)

Requirement already satisfied: matplotlib!=3.6.1,>=3.4 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from seaborn) (3.8.4)

Requirement already satisfied: contourpy>=1.0.1 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.2.1)

Requirement already satisfied: cycler>=0.10 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (0.12.1)

Requirement already satisfied: fonttools>=4.22.0 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (4.51.0)

Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (1.4.5)

Requirement already satisfied: packaging>=20.0 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (24.0)

Requirement already satisfied: pillow>=8 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (10.3.0)

Requirement already satisfied: pyparsing>=2.3.1 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (3.1.2)

Requirement already satisfied: python-dateutil>=2.7 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from matplotlib!=3.6.1,>=3.4->seaborn) (2.9.0.post0)

Requirement already satisfied: pytz>=2020.1 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from pandas>=1.2->seaborn) (2024.1)

Requirement already satisfied: tzdata>=2022.7 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from pandas>=1.2->seaborn) (2024.1)

Requirement already satisfied: six>=1.5 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.4->seaborn) (1.16.0)

Note: you may need to restart the kernel to use updated packages.

In [3]: `pip install scikit-learn`

Requirement already satisfied: scikit-learn in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (1.4.2)

Requirement already satisfied: numpy>=1.19.5 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from scikit-learn) (1.26.4)

Requirement already satisfied: scipy>=1.6.0 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from scikit-learn) (1.13.0)

Requirement already satisfied: joblib>=1.2.0 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from scikit-learn) (1.4.2)

Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from scikit-learn) (3.5.0)

Note: you may need to restart the kernel to use updated packages.

In [4]: `pip install librosa`

Requirement already satisfied: librosa in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (0.10.2)

Requirement already satisfied: audioread>=2.1.9 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from librosa) (3.0.1)

Requirement already satisfied: numpy!=1.22.0,!=1.22.1,!=1.22.2,>=1.20.3 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from librosa) (1.26.4)

Requirement already satisfied: scipy>=1.2.0 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from librosa) (1.13.0)

Requirement already satisfied: scikit-learn>=0.20.0 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from librosa) (1.4.2)

Requirement already satisfied: joblib>=0.14 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from librosa) (1.4.2)

Requirement already satisfied: decorator>=4.3.0 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from librosa) (5.1.1)

Requirement already satisfied: numba>=0.51.0 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from librosa) (0.59.1)

Requirement already satisfied: soundfile>=0.12.1 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from librosa) (0.12.1)

Requirement already satisfied: pooch>=1.1 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from librosa) (1.8.1)

Requirement already satisfied: soxr>=0.3.2 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from librosa) (0.3.7)

Requirement already satisfied: typing-extensions>=4.1.1 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from librosa) (4.11.0)

Requirement already satisfied: lazy-loader>=0.1 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from librosa) (0.4)

Requirement already satisfied: msgpack>=1.0 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from librosa) (1.0.8)

Requirement already satisfied: packaging in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from lazy-loader>=0.1->librosa) (24.0)

Requirement already satisfied: llvmlite<0.43,>=0.42.0dev0 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from numba>=0.51.0->librosa) (0.42.0)

Requirement already satisfied: platformdirs>=2.5.0 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from pooch>=1.1->librosa) (4.2.0)

Requirement already satisfied: requests>=2.19.0 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from pooch>=1.1->librosa) (2.31.0)

Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from scikit-learn>=0.20.0->librosa) (3.5.0)

Requirement already satisfied: cffi>=1.0 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from soundfile>=0.12.1->librosa) (1.16.0)

Requirement already satisfied: pycparser in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from cffi>=1.0->soundfile>=0.12.1->librosa) (2.22)

Requirement already satisfied: charset-normalizer<4,>=2 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from requests>=2.19.0->pooch>=1.1->librosa) (3.3.2)

Requirement already satisfied: idna<4,>=2.5 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from requests>=2.19.0->pooch>=1.1->librosa) (3.7)

Requirement already satisfied: urllib3<3,>=1.21.1 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from requests>=2.19.0->pooch>=1.1->librosa) (2.2.1)

Requirement already satisfied: certifi>=2017.4.17 in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from requests>=2.19.0->pooch>=1.1->librosa) (2024.2.2)

Note: you may need to restart the kernel to use updated packages.

```
In [5]: # Usual Libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
import sklearn

# Librosa (the mother of audio files)
import librosa
import librosa.display
import IPython.display as ipd
import warnings
warnings.filterwarnings('ignore')
```

```
In [6]: import os
general_path = "C:\\Users\\Vasanth Soorya\\Downloads\\archive (1)\\Data"
print(list(os.listdir(f'{general_path}/genres_original/')))
```

```
['blues', 'classical', 'country', 'disco', 'hiphop', 'jazz', 'metal', 'pop', 'reggae', 'rock']
```

```
In [7]: # Importing 1 file
y, sr = librosa.load(f'{general_path}/genres_original/reggae/reggae.00036.wav')

print('y:', y, '\n')
print('y shape:', np.shape(y), '\n')
print('Sample Rate (KHz):', sr, '\n')

# Verify Length of the audio
print('Check Len of Audio:', 661794/22050)
```

```
y: [0.02072144 0.04492188 0.05422974 ... 0.06912231 0.08303833 0.08572388]
```

```
y shape: (661794,)
```

```
Sample Rate (KHz): 22050
```

```
Check Len of Audio: 30.013333333333332
```

```
In [8]: # Trim leading and trailing silence from an audio signal (silence before and after
audio_file, _ = librosa.effects.trim(y)

# the result is an numpy ndarray
print('Audio File:', audio_file, '\n')
print('Audio File shape:', np.shape(audio_file))
```

```
Audio File: [0.02072144 0.04492188 0.05422974 ... 0.06912231 0.08303833 0.08572388]
```

```
Audio File shape: (661794,)
```

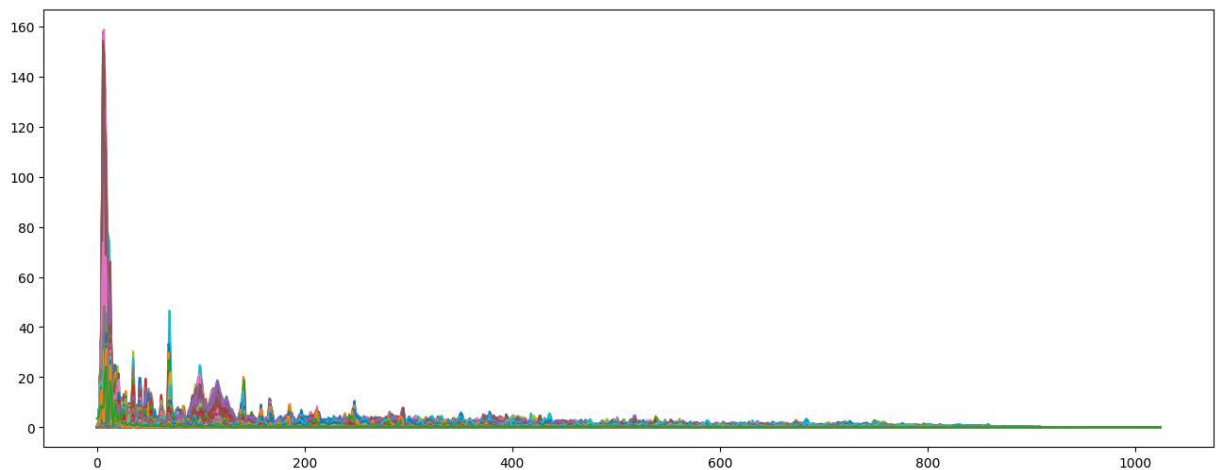
```
In [9]: # Default FFT window size
n_fft = 2048 # FFT window size
hop_length = 512 # number audio of frames between STFT columns (Looks Like a good d

# Short-time Fourier transform (STFT)
D = np.abs(librosa.stft(audio_file, n_fft = n_fft, hop_length = hop_length))

print('Shape of D object:', np.shape(D))
```

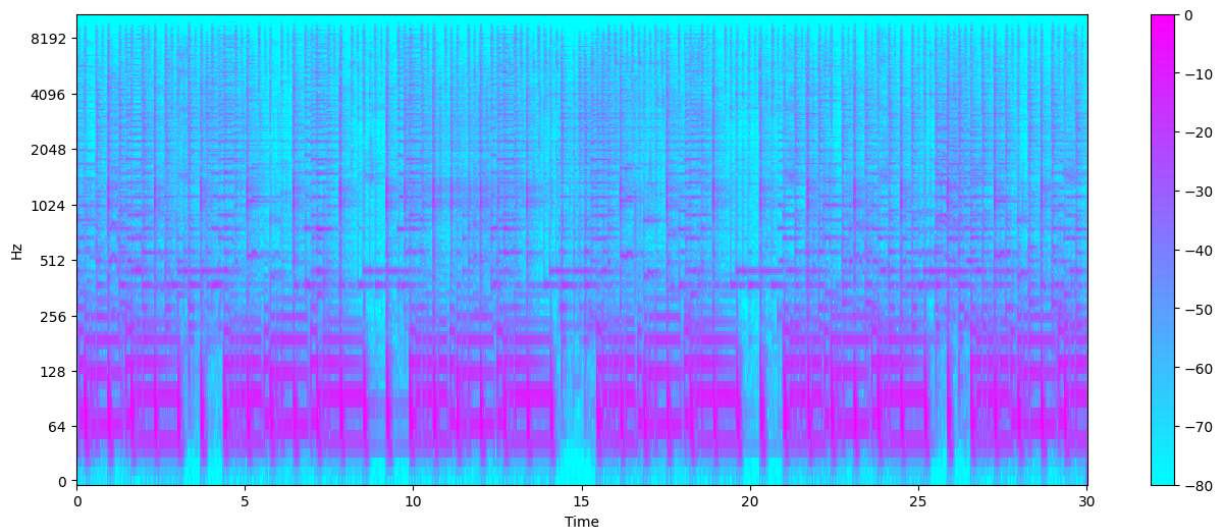
Shape of D object: (1025, 1293)

```
In [10]: plt.figure(figsize = (16, 6))
plt.plot(D);
```



```
In [11]: # Convert an amplitude spectrogram to Decibels-scaled spectrogram.
DB = librosa.amplitude_to_db(D, ref = np.max)

# Creating the Spectrogram
plt.figure(figsize = (16, 6))
librosa.display.specshow(DB, sr = sr, hop_length = hop_length, x_axis = 'time', y_a
                        cmap = 'cool')
plt.colorbar();
```

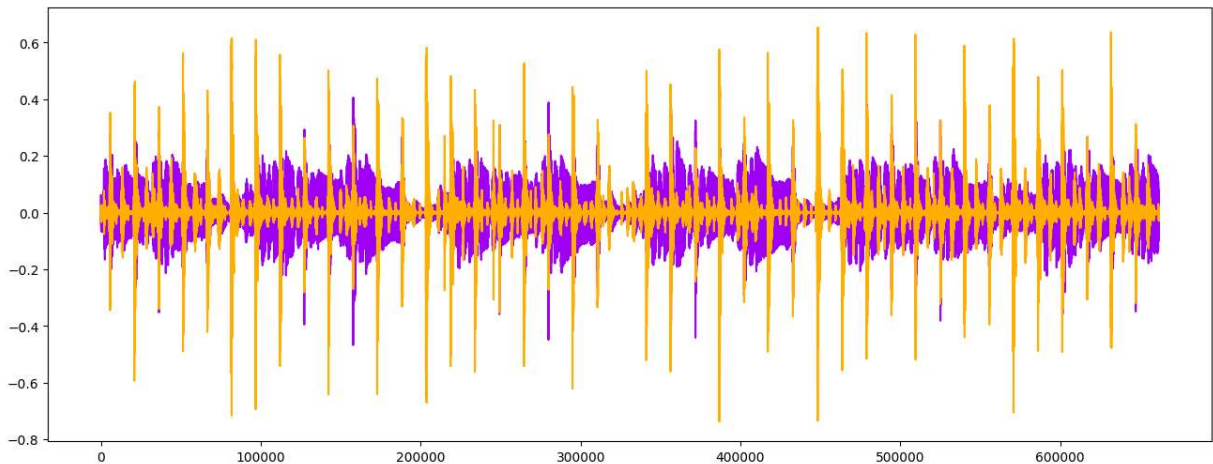


```
In [12]: # Total zero_crossings in our 1 song
zero_crossings = librosa.zero_crossings(audio_file, pad=False)
print(sum(zero_crossings))
```

39232

```
In [13]: y_harm, y_perc = librosa.effects.hpss(audio_file)

plt.figure(figsize = (16, 6))
plt.plot(y_harm, color = '#A300F9');
plt.plot(y_perc, color = '#FFB100');
```



```
In [14]: pip install xgboost
```

Requirement already satisfied: xgboost in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (2.0.3)
 Requirement already satisfied: numpy in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from xgboost) (1.26.4)
 Requirement already satisfied: scipy in c:\users\vasanth soorya\appdata\local\programs\python\python311\lib\site-packages (from xgboost) (1.13.0)
 Note: you may need to restart the kernel to use updated packages.

```
In [15]: from sklearn.naive_bayes import GaussianNB
from sklearn.linear_model import SGDClassifier, LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.neural_network import MLPClassifier
from xgboost import XGBClassifier, XGBRFClassifier
from xgboost import plot_tree, plot_importance

from sklearn.metrics import confusion_matrix, accuracy_score, roc_auc_score, roc_curve
from sklearn import preprocessing
from sklearn.model_selection import train_test_split
from sklearn.feature_selection import RFE
```

```
In [16]: data = pd.read_csv(f'{general_path}/features_3_sec.csv')
data = data.iloc[0:, 1:]
data.head()
```

```
Out[16]:
```

	length	chroma_stft_mean	chroma_stft_var	rms_mean	rms_var	spectral_centroid_mean
0	66149	0.335406	0.091048	0.130405	0.003521	1773.065032
1	66149	0.343065	0.086147	0.112699	0.001450	1816.693777
2	66149	0.346815	0.092243	0.132003	0.004620	1788.539719
3	66149	0.363639	0.086856	0.132565	0.002448	1655.289045
4	66149	0.335579	0.088129	0.143289	0.001701	1630.656199

5 rows × 59 columns



```
In [17]: y = data['label'] # genre variable.
X = data.loc[:, data.columns != 'label'] #select all columns but not the labels

#### NORMALIZE X ####

# Normalize so everything is on the same scale.

cols = X.columns
min_max_scaler = preprocessing.MinMaxScaler()
np_scaled = min_max_scaler.fit_transform(X)

# new data frame with the new scaled data.
X = pd.DataFrame(np_scaled, columns = cols)
```

```
In [18]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_sta
```

```
In [19]: def model_assess(model, title = "Default"):
    model.fit(X_train, y_train)
    preds = model.predict(X_test)
    #print(confusion_matrix(y_test, preds))
    print('Accuracy', title, ':', round(accuracy_score(y_test, preds), 5), '\n')
```

```
In [20]: # Naive Bayes
nb = GaussianNB()
model_assess(nb, "Naive Bayes")

# Stochastic Gradient Descent
sgd = SGDClassifier(max_iter=5000, random_state=0)
model_assess(sgd, "Stochastic Gradient Descent")

# KNN
knn = KNeighborsClassifier(n_neighbors=19)
model_assess(knn, "KNN")

# Decission trees
tree = DecisionTreeClassifier()
model_assess(tree, "Decission trees")

# Random Forest
```

```

rforest = RandomForestClassifier(n_estimators=1000, max_depth=10, random_state=0)
model_assess(rforest, "Random Forest")

# Support Vector Machine
svm = SVC(decision_function_shape="ovo")
model_assess(svm, "Support Vector Machine")

# Logistic Regression
lg = LogisticRegression(random_state=0, solver='lbfgs', multi_class='multinomial')
model_assess(lg, "Logistic Regression")

# Neural Nets
nn = MLPClassifier(solver='lbfgs', alpha=1e-5, hidden_layer_sizes=(5000, 10), random_state=0)
model_assess(nn, "Neural Nets")

```

Accuracy Naive Bayes : 0.51952

Accuracy Stochastic Gradient Descent : 0.65532

Accuracy KNN : 0.80581

Accuracy Decision trees : 0.6343

Accuracy Random Forest : 0.81415

Accuracy Support Vector Machine : 0.75409

Accuracy Logistic Regression : 0.6977

Accuracy Neural Nets : 0.67134

```

In [21]: # Libraries
import IPython.display as ipd
from sklearn.metrics.pairwise import cosine_similarity
from sklearn import preprocessing

# Read data
data = pd.read_csv(f'{general_path}/features_30_sec.csv', index_col='filename')

# Extract labels
labels = data[['label']]

# Drop labels from original dataframe
data = data.drop(columns=['length', 'label'])
data.head()

# Scale the data
data_scaled=preprocessing.scale(data)
print('Scaled data type:', type(data_scaled))

```

Scaled data type: <class 'numpy.ndarray'>

```

In [22]: # Cosine similarity
similarity = cosine_similarity(data_scaled)
print("Similarity shape:", similarity.shape)

```

```
# Convert into a dataframe and then set the row index and column names as labels
sim_df_labels = pd.DataFrame(similarity)
sim_df_names = sim_df_labels.set_index(labels.index)
sim_df_names.columns = labels.index

sim_df_names.head()
```

Similarity shape: (1000, 1000)

Out[22]:

filename	blues.00000.wav	blues.00001.wav	blues.00002.wav	blues.00003.wav	blu
blues.00000.wav	1.000000	0.049231	0.589618	0.284862	
blues.00001.wav	0.049231	1.000000	-0.096834	0.520903	
blues.00002.wav	0.589618	-0.096834	1.000000	0.210411	
blues.00003.wav	0.284862	0.520903	0.210411	1.000000	
blues.00004.wav	0.025561	0.080749	0.400266	0.126437	

filename

blues.00000.wav	1.000000	0.049231	0.589618	0.284862
blues.00001.wav	0.049231	1.000000	-0.096834	0.520903
blues.00002.wav	0.589618	-0.096834	1.000000	0.210411
blues.00003.wav	0.284862	0.520903	0.210411	1.000000
blues.00004.wav	0.025561	0.080749	0.400266	0.126437

5 rows × 1000 columns



In [23]:

```
def find_similar_songs(name):
    # Find songs most similar to another song
    series = sim_df_names[name].sort_values(ascending = False)

    # Remove cosine similarity == 1 (songs will always have the best match with the
    series = series.drop(name)

    # Display the 5 top matches
    print("\n*****\nSimilar songs to ", name)
    print(series.head(5))
```

In [26]: `ipd.Audio(f'{general_path}/genres_original/pop/pop.00023.wav')`

Out[26]:

In [27]: `ipd.Audio(f'{general_path}/genres_original/pop/pop.00034.wav')`

Out[27]:

In [28]:

```
# metal.00002 - Iron Maiden "Flight of Icarus"
find_similar_songs('metal.00002.wav')

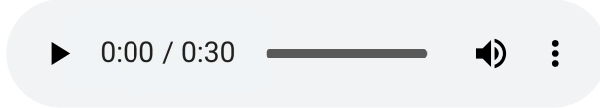
ipd.Audio(f'{general_path}/genres_original/metal/metal.00002.wav')
```

Similar songs to metal.00002.wav
filename

metal.00028.wav	0.904367
metal.00059.wav	0.896096
rock.00018.wav	0.891910
rock.00017.wav	0.886526
rock.00016.wav	0.867508

Name: metal.00002.wav, dtype: float64

Out[28]:



In [29]:

```
ipd.Audio(f'{general_path}/genres_original/rock/rock.00016.wav')
```

Out[29]:

