Andrew Hill, ahill6@ncsu.edu
Dustin Lambright, dalambri@ncsu.edu
Fuxing Luan, fluan@ncsu.edu
Guanxu Yu, gyu9@ncsu.edu
Marc Quaintance, miquaint@ncsu.edu
Yuchen Sun, ysun34@ncsu.edu

PROJECT: ALGO-RHYTHM
Team Name: **Skinny White Godzillas, Thick Grunting Diggers**
Unemployed Wheelers, Old Driver

Overview
    Language(s) we are using
        We will be using Java and Python
    Data set we are using
        http://labrosa.ee.columbia.edu/millionsong/
        http://labrosa.ee.columbia.edu/millionsong/pages/field-list
    Goal
        Our goal is to classify songs with the focus of predicting if a song will make it to the top 40

Question(s) to Answer
    3 easy questions to validate our setup
        - Can we parse the .h5 files?
        - Can we pick out the parts that we *think* will help us out the most?
        - Can we attach song success/genre to the already existing songs?

    2-6 medium questions we know we can do
        - Build a classification model
        - Cluster the songs
        - Instructions for how to write a hit song
        - Show the trends of popular music

    2-3 stretch goals that we may not even finish, but would be awesome if we did
        - Calculate the genre of a song (Rock, Pop, Country, etc) based on data points
        - Predict the number of weeks a song will be on the top 40
        - Recommender system for a user (based on user's favorite genres, artists)

Plan - What Will We Use To Answer Questions

Preprocessing
        We will most likely use Java to parse the .h5 files since it has a built-in functionality to

parse the .h5 formats. We will *most likely* remove the unstructured data, and export our results as JSON or .csv files.

Clustering Algorithms (if any)
>	k-means

Classification Algorithms (if any)
>	Convolutional neural network with TensorFlow, other algorithms we learn in class

Outlier Detectors (if any)
>	Principal components analysis

Postprocessing
>	Use data visualization techniques to display the result more interactively

Division of Labor

Section	Task	Preferred Job	Assigned To
Preprocessing			
	I/O	Dustin	Dustin
	Data Cleaning	Yuchen	
Clustering			
	Method 1	Yuchen	
	Method 2	Dustin	
Classification			
	Method 1	Guanxu	
	Method 2	Marc	
	Method 3	Fuxing	
	Method 4(?)	Andrew	
Outlier Detection			
	Method 1(?)	Andrew	
Postprocessing			
	Results cleanup – clustering	Guanxu	
	Results cleanup – classification	Marc	

	Results cleanup – outliers(?)	Fuxing	
	Analysis of results	Guanxu	
Visualization			
	Making graphs/charts	Yuchen	
	Making pretty pictures	Andrew	
Responsibilities for Final Write-up			
	Who must do what	Marc	

Articles that we could use:

http://www.sciencedirect.com/science/article/pii/S1532046403000340
http://www.ee.columbia.edu/~dliang/files/FINAL.pdf
https://pdfs.semanticscholar.org/6ca7/b53455585b484afef120cfea75ebfc84f7d6.pdf