# CS 4780 Final Project Proposal: What Makes a Song "Beatlesque"?

## 1. Size of Team

There are four (4) members in our team.

## 2. Motivation

The Beatles are considered one of the most influential bands of all time. We think it will be an interesting and challenging task to formalize how that influence is manifested in the metadata of the work of other artists.

## 3. Problem Statement

It is difficult to formalize what being "influenced by The Beatles" means, so our main learning task is to determine the relationship between a song's metadata and the extent that the song (and by extension, its artist) is influenced by the work of The Beatles.

More concisely, we want to determine how similarity in song metadata translates to musical influence.

## 4. General Approach

### 4.1. Background Knowledge

This will be a supervised learning problem, so we need to acquire background information on different ways musical influence manifests itself in a song.

The Beatles have a large corpus of music (over 240 songs), so we will study the 25 most popular songs by The Beatles in our training, testing, and analysis.

### 4.2. Data Parsing and Clean Up

The Million Song Dataset includes attributes unnecessary for our goals, so we will only focus on certain features such as artist familiarity, tags describing the artist, energy, year, and similar artists. We are open to using more attributes and will learn more important ones in our main learning exercise.

### 4.3. Learning Task Setup

For our first round of training, we will use 200 sample songs - 50 of which we have already determined as "influenced" by the Beatles, in order to develop better intuition for the nature of The Beatles' influence and begin understanding how to compute weights in the features before officially training and testing on our remaining data. An additional 1,000 songs will be used for second-round training and validation. 1,000 songs will be used for testing.

### 4.4. Training

We will test various similarity measures on small subsets of our data to learn what features are most important in determining The Beatles' influence on a song.

Once we find a strong similarity measure, we will implement the K Nearest Neighbor algorithm to provide a classification of an artist given the metadata and similarity of songs by that artist to The Beatles.

## 5. Resources

Dataset: Million Song Dataset, Columbia University.
Language: Matlab, useful data visualization packages.
Version Control: git, private repository on GitHub.

## 6. Schedule

| Date | Milestone |
|---|---|
| **10/21** | **Project Proposal Due** |
| 10/26 | Finish Background Research |
| 11/7 | Data Parsing* |
| 11/9 | Influence Linked to Song Attributes* |
| **11/11** | **Progress Report Due** |
| 11/15 | Testing of Various Similarity Measures* |
| 11/25 | Implementation of KNN & Testing |
| 11/28 | Evaluation of Findings |
| 12/02 | Finish Final Report** |
| 12/02 | Finish Poster** |
| **12/04** | **Poster Presentation** |
| **12/16** | **Final Project Report Due** |

*, ** Tasks occurring concurrently
**Bold face text** Hard deadline from course instructors