**Project Proposal**

***Team:*** Sabrina Shen (No Partners)

**Problem:**

For this project, I will be trying to tackle instrument recognition through machine learning. I will start by defining the goal of the algorithm to be recognizing the instrument in an audio track that only contains music produced by a single instrument; if this initial algorithm performs well, I would like to expand the project to detect multiple instruments within a single audio track.

**Dataset:**

The project restrictions state “You must use at least one dataset with at least one half million data points as a significant part of your project”; however, I have been unable to find a single dataset with more than 500k audio samples, below I have instead listed largest datasets I had found:

* Single instrument audio
  + NSynth Dataset (305,979)
  + AudioSet Musical instrument (117,343, unsure if accessible)
  + IRMAS (6,705)
  + UIOWA MIS (2,182)
  + RWC (3,544)
  + Good-sounds.org dataset (cannot validate size without downloading but seems robust from the description)
* Multi-instrument audio
  + MeledyDB (122)
  + GTZAN Dataset (1,000)

**Methods:**

My current approach to this problem is to first preprocess the dataset and extract audio features that will be used to train the model. From my previous work in audio signal processing I believe that the spectrogram, chroma features, and MFCC will be useful in identifying unique features of each instrument.

I have a few algorithms that I am currently considering for the model architecture. I believe that algorithms such as Random Decision Forest (although I don’t know if it is technically covered in this course), Naïve Bayes, Support Vector Machines, and Principal component analysis could all perform well and they are easily implementable with python libraries so the methods would be simple to check. I am also considering approaching this problem by implementing and extending the NMF algorithm. I believe that by taking this approach, I may be able to not only identify a single instrument but also, by ensembling the activation matrices for each instrument, identify multiple instruments within a single audio track.