Clark Brown, Sam Cochran, Daniel Swingle

Data Scraped and Cleaned:

- Our data is very uniform in the way names are spelled. We have 8 genres (all spelled/labelled uniformly) and a filename for each audio sample. The filenames are just numbers (in string form, of course), so those are very uniform and easy to sample from.
- The data has been uniformized with regard to format—we've converted each audio sample to a .wav file format.
- During our processing of each mp3 files to wav files, we found that 3 tracks did not convert. Because the dataset is composed of 8000 files, this should not be a large issue, but it still is something to keep in mind as far as balancing genres.
- Now that the data is scraped and cleaned, we have a solid set to work with. The next challenge will be extracting features and finding which features will be the most important for classifying our audio files—this will be the hardest part of the project. But our data is currently well set up to be able to move on to this next stage.

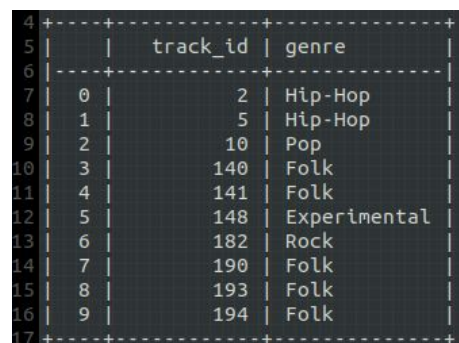Evaluating the Reliability of the Source, Strengths and Weaknesses of the Dataset

- One potential issue with our data is that the dataset is composed entirely of free music, and therefore our model may have difficulty analyzing other kinds of music, which may be quite different.
- Specifically, we have reason to believe that the genre definitions, quality, and style of a free music database may differ from commercial music, so we will have to find a way to evaluate how well a model trained on a free music database can generalize to samples of commercial music.

Revise our plan if Necessary

- Overall, things have gone pretty much according to plan—we did a fair amount of data analysis when choosing this dataset in particular, and we should be able to proceed as planned in our proposal.

Dataset Description

- The source is The Free Music Archive created by Michaël Defferrard, Kirell Benzi, Pierre Vandergheynst, Xavier Bresson. International Society for Music Information Retrieval Conference (ISMIR), 2017.
- 8 genres: 'Hip-Hop', 'Pop', 'Folk', 'Experimental', 'Rock', 'International', 'Electronic', 'Instrumental'
- 8000 tracks evenly distributed across genres (1000 tracks per genre)

```
 4 +----+------------+--------------+
 5 |    |  track_id  | genre        |
 6 |----+------------+--------------|
 7 | 0  |          2 | Hip-Hop      |
 8 | 1  |          5 | Hip-Hop      |
 9 | 2  |         10 | Pop          |
10 | 3  |        140 | Folk         |
11 | 4  |        141 | Folk         |
12 | 5  |        148 | Experimental |
13 | 6  |        182 | Rock         |
14 | 7  |        190 | Folk         |
15 | 8  |        193 | Folk         |
16 | 9  |        194 | Folk         |
17 +----+------------+--------------+
```