

The dataset for our project consists of music information combined from various sources such as

1. Million song dataset(msd)
2. Last.fm dataset
3. Msd genre dataset
4. Msd user data set

#### 1. Million song dataset

This dataset consists of various files in .hdf5 format where each file corresponds to single track. Each file consists of track information such as title, trackId, duration, loudness and year the track released.

Also it consists of song information, its name and id. Also it consists of Album information, its name and release year.

The artist data is also included in each file. The data consists of artistId, location, name, various terms related to artists and list of artists which are similar to it.

#### 2. Last.fm dataset

This dataset consists of various files in json format. Each file consists of trackIds of tracks that are similar to given track and its confidence value indicating how much similar it is. Also it consists of various tags for a particular track.

#### 3. MSD genre dataset

This dataset consists of genre information of various tracks from MSD dataset. So was combined with MSD dataset by using trackID

#### 4. MSD user data set

This dataset consists of data about the listening users. The data consists of the the userId, the SongID of songs they listen and number of times song is played.