#### Collection

* 14,592 tracks with 239,620 comments of which 237,196 are timed comments collected based upon most recent upload date ranging from Jan 1, 2010 to Jan 1, 2014.
* Track data here -<https://developers.soundcloud.com/docs/api/reference#tracks>
* Comment data here - <https://developers.soundcloud.com/docs/api/reference#comments>
* Categorized based on 5 licenses sub-categorized based on 36 genres.
* Tracks with at least one comment are taken into consideration.
* Stored in pickle files which comprise of serialized objects.
* Only the following licenses were allowed through the API – ‘cc-by’, ‘cc-by-sa’, ‘cc-by-nd’, ‘cc-by-nc-nd’, ‘cc-by-nc-sa’. Others returned HTTP 400 Bad request when explicitly passing a license argument.

#### Issues

* Genre of track does not necessarily imply the genre as referenced by the API, thereby leads to duplicates in other genres and discrepancies.
* Genre of track is user defined therefore categorizing based upon this ever growing field would prove tedious.
* Track.comment\_count (245,016) does not necessarily imply the actual comment count due to special cases such as <https://soundcloud.com/djraisdnet/3hr-dj-set>

#### Considerations

* Look for comment distribution in genre.
* Consider playback, download and favorites count.
* Study user interest in genres.
* Consider the life time of a track.

#### Filtering

* Tracks with less than or equal to 3 were removed.
* Tracks which were only streamable considered.

#### Algorithms

* Adjacent (Martha’s) – (time\_end\_comment– time\_start\_comment)/max(time\_between\_adj)
* First week (Karthik’s) – (total\_no\_comments)/(no\_comments\_first\_week)
* Relative week (Mine):  
    
  score1 = (no\_comments\_week) >= (no\_comments\_week -1 \* 0.5) [Note : not consecutive weeks eg:- some tracks had following week activity : 20, 1, 15 in the list of no\_comments in each week]  
    
  score2 = (no\_comments\_week)\*(i) [‘i’ signifies iteration]