#### Collection

* 36,180 tracks with around 100k 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

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

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.
* Special cases such as <https://soundcloud.com/djraisdnet/3hr-dj-set>

#### Proposal

* Pick 3 tracks from each genre – most, mediocre and least popular.
* Filter out spam and non-english comments.
* Ensure that the 100 tracks comprise of negative comments as well.

#### Refining

* A simple formula for popularity rating favoritings\_count\*0.6 + download\_count\*0.4 was used to sort tracks.
* In the sorted list for each genre the second highest and second lowest popular track were selected.