**"Introduction to Data Science" project – Ben Lahav & Ben Bukai**

Research question approval for final project

1. Research questions

Can we predict if a given song has inappropriate words in it?

1. Fundamental project overview and milestones:
   1. Data Acquisition
2. APIs
   1. Spotify API to get thousands of tracks

<https://developer.spotify.com/>

* 1. Lyrics API for songs

<https://rapidapi.com/collection/lyrics-apis>

* 1. Inappropriate API decider API

<https://promptapi.com/marketplace/description/bad_words-api>

1. Crawling
   1. Trying to get a hold of extra data about artist's mentality or special violence lawsuit that has effect on the probability of inappropriate words in his/hers song.
2. Extra features self developed
   1. Maybe we'll do aggregate operations on each feature that has info about bad words that we are developing. For instance we'll create a feature "past bad words probability" or counting for an artist.
   2. Data Cleaning
3. Using data mining process we would clean the data with the pipeline of Data cleaning techniques as we learned through out the course.
   1. EDA
4. Visualize what we've got so far
   * 1. Amount of songs with bad word by relation to number of songs.
     2. Amout of songs per period.
     3. Etc.
   1. ML
5. Finding best ML approach and algorithm to this problem.
6. Hopefully this tool would provide parents or businesses a better knowing about probability of violent content in artists creation by their past.
7. Show a couple of ML algorithms predictions and success percentages.