

Exploring the Sentiment of Lyrics from Billboard Top 100 1960-2015

Tuomo Kareoja

December 17, 2019

Agenda

Getting the Data

Data Cleaning

Measuring Sentiment

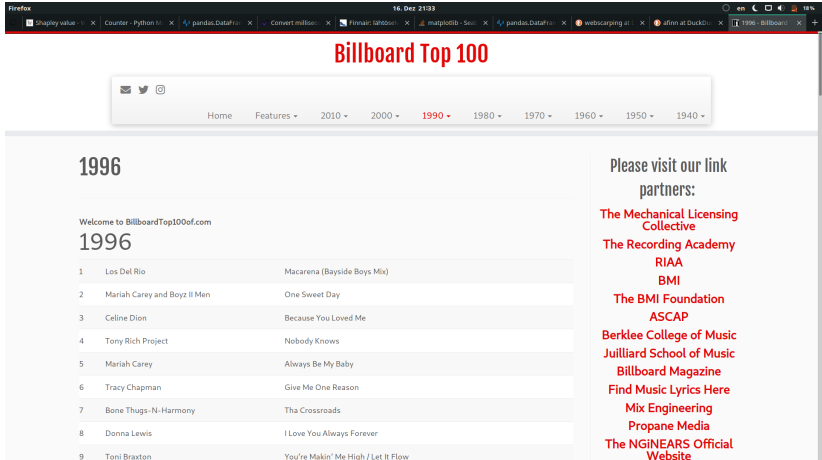
Results

- Audio Features

- Lyrics Sentiment

Conclusion

- ▶ Reading the contents from HTML tables with the BeautifulSoup package



The screenshot shows a web browser window with the URL "1996 - Billboard Top 100". The page title is "Billboard Top 100". The navigation bar includes links for Home, Features, and a year selector with options from 2010 to 1940, with 1990 selected. The main content area displays the year "1996" and a table of the top 10 songs. To the right, there is a section titled "Please visit our link partners:" listing various music-related organizations and websites.

Rank	Artist	Song
1	Los Del Rio	Macarena (Bayside Boys Mix)
2	Mariah Carey and Boyz II Men	One Sweet Day
3	Celine Dion	Because You Loved Me
4	Tony Rich Project	Nobody Knows
5	Mariah Carey	Always Be My Baby
6	Tracy Chapman	Give Me One Reason
7	Bone Thugs-N-Harmony	Tha Crossroads
8	Donna Lewis	I Love You Always Forever
9	Toni Braxton	You're Makin' Me High / Let It Flow

Please visit our link partners:

- The Mechanical Licensing Collective
- The Recording Academy
- RIAA
- BMI
- The BMI Foundation
- ASCAP
- Berklee College of Music
- Juilliard School of Music
- Billboard Magazine
- Find Music Lyrics Here
- Mix Engineering
- Propane Media
- The NGINEARS Official Website

Lyrics From APIs

- ▶ Used three different APIs with the Requests package
 1. Mourits Lyrics (best, but so much down!)
 2. Genius Lyrics
 3. ChartLyrics
- ▶ If the lyrics is too long, try a different API (no long bible quotes!)
- ▶ What you get differs from site to site as the lyrics are added by users

Get Track Details From Spotify

- ▶ Spotify has a really nice API for getting information about the actual audio (e.g. length, danceability, tempo, loudness...)
- ▶ You need to register an app, and the API, is bit fiddly...
- ▶ But of course there is a python package does most of the work for you! (spotipy)
- ▶ Still need to first find the song ID and then find the audio features with this, so multiple calls needed

Cleaning the Lyrics

- ▶ Lyrics are typed by humans so they contain typos and trolling

Rihanna - You Da One

Hello! dafdcdk interesting dafdcdk site! I'm really like it! Very, very dafdcdk good!

- ▶ Non-standard spelling is commonplace

Dr. Dre - Nothing But a G Thang

...yo' attention Mobbin like a muh'fucker, but I ain't lynchin Droppin the funky shit that's makin the sucka niggaz mumble...

- ▶ And so many other problems too...

Is It Instrumental?

- ▶ Try to find mention of song being instrumental from lyrics
 - ▶ "Instrumental, NO LYRICS"
 - ▶ What if there is an instrumental part marked in the lyrics?
- ▶ Very short lyrics
 - ▶ "Ah ha ha ha ha ha ha ha ha ha ha, wipe out"
- ▶ Use Spotify instrumentality
 - ▶ Gives an estimated probability of the song being instrumental

Is It In English?

- ▶ Billboard contains also some Spanish, French and Italian(!) songs
- ▶ These can be recognized with the langdetect package
- ▶ ...but some songs just have spanish parts (or trolling or mistakes)

Rod Stewart - Do You Think I'm Sexy

...Cariño, cariño Ella se sienta sola esperando propuestas Es está muy nervioso evitando todas las preguntas Los labios de él están secos...

Measuring Sentiment

Afinn

- ▶ Lexicon of English terms manually rated for valence with an integer between -5 (negative) and 5 (positive)
 - ▶ `afinn.score('Terrible') = -3.0`
 - ▶ `afinn.score('Make Berlin great again!') = 3.0`

Emolex

- ▶ Lexicon of English words manually rated for associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive)

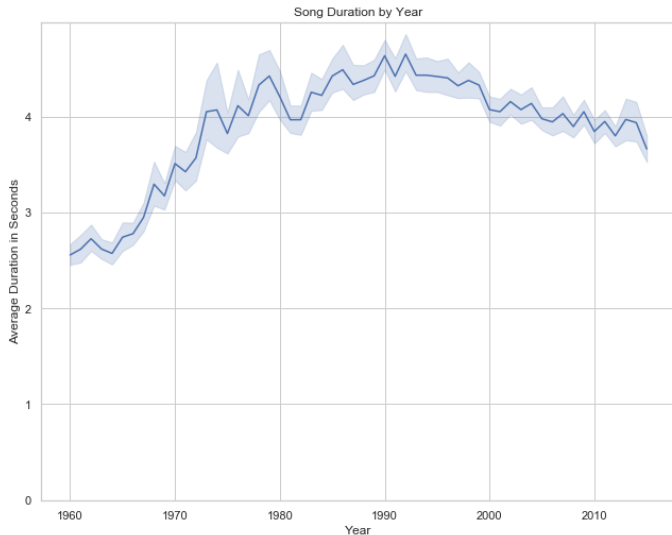
abnormal disgust 1

abnormal negative 1

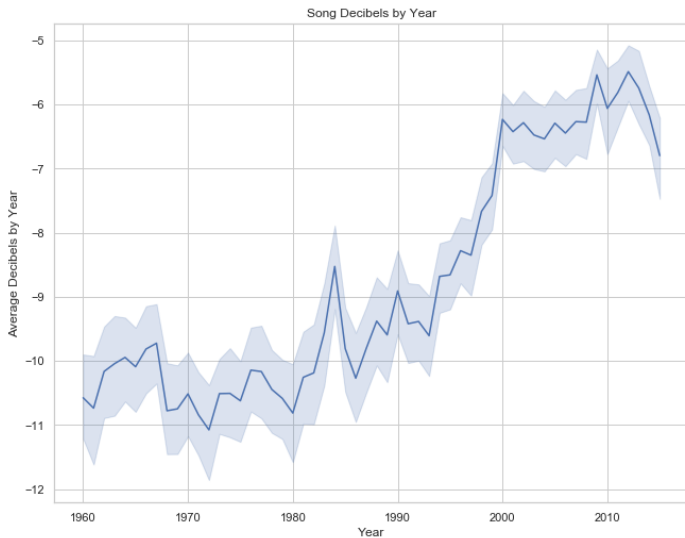
Results

Audio Features

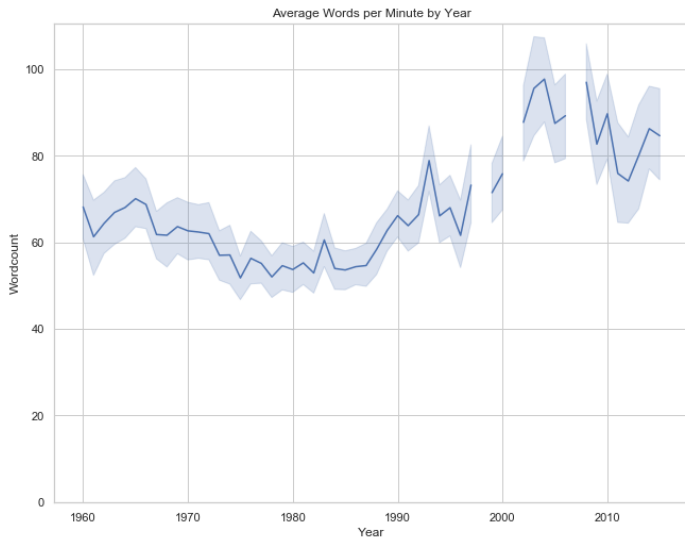
Technology Has Pushed Song Duration Up, But Not Anymore



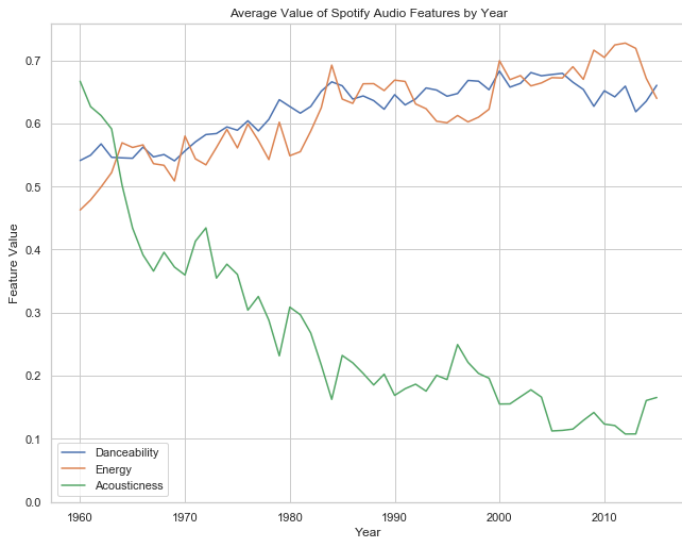
Songs are Getting Louder



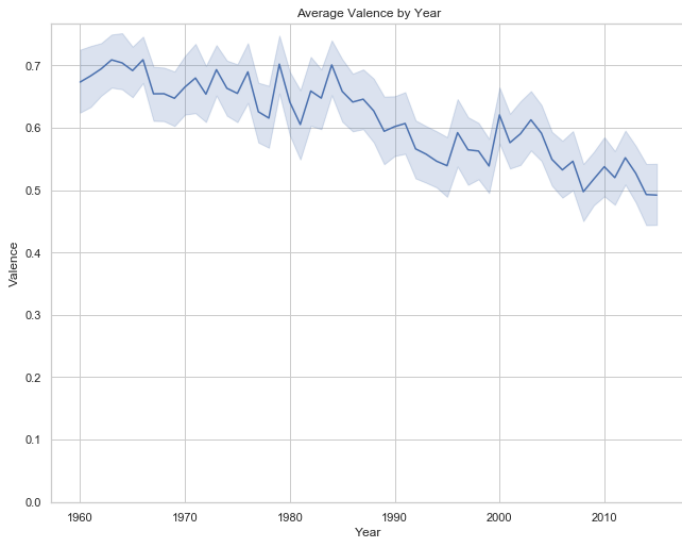
More Words per Minute (Rap's Influence?)



Death of Acousticness/Rise of Dance Music

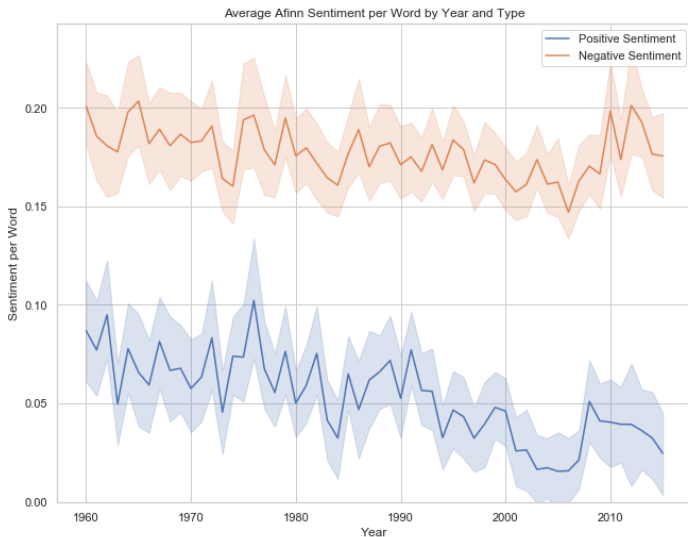


Audio Features Show Decreasing Positivity

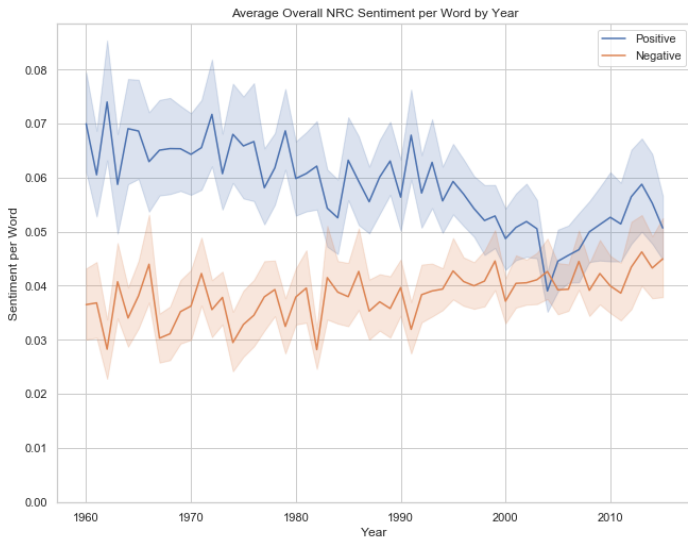


Lyrics Sentiment

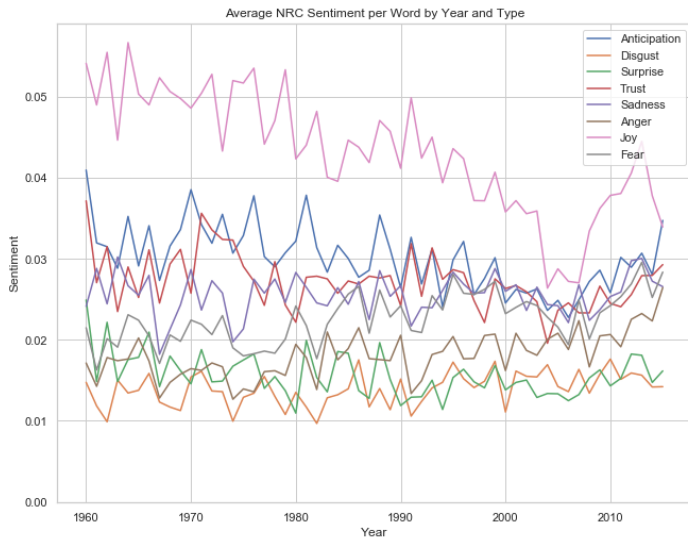
Afinn Positive Sentiment is in Decline



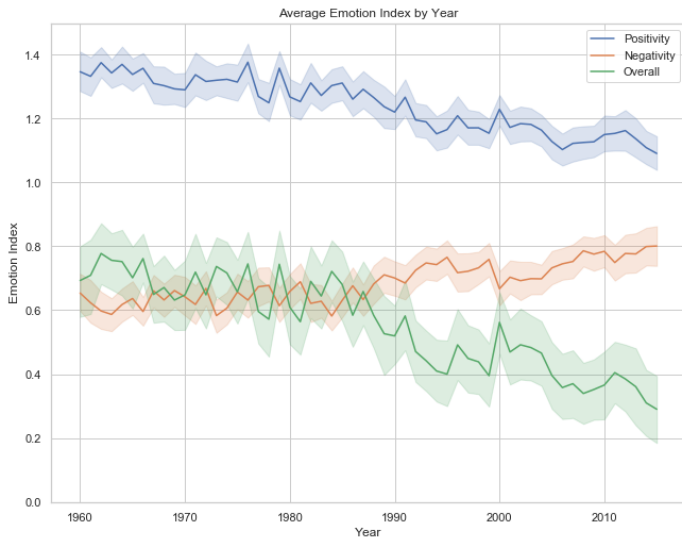
NRC Shows the Same



This Is Because There Is Less Joy In Lyrics



Indexed Sentiment Goes Down



Conclusions

Hits have become:

1. Longer
2. Louder
3. More danceable
4. Less happy (just from audio features)

Lyrics of the hits have become:

1. Longer
2. Less positive (less joy and positiveness)

- ▶ Most positive song: Hank Ballard and The Midnighters - Finger Poppin' Time, 1960
- ▶ Most negative song: Tyga - Rack City, 2012
- ▶ Most words: Eminem - Rap God, 2013
- ▶ Most unique words: Eminem - Rap God, 2013

The End

Questions?