Sentiment Analysis of 50 Years Worth of Top Billboard Songs

Maggie Tjia

Columbia University

**Abstract**

The reception of music has been wide and universal. Music is an art form that almost everyone utilizes at one point or another. The current project looks at the change in music that has made it to the Billboard charts throughout 50 years in terms of the sentiment of the music’s lyrics (if any). The motivation was to see whether music has changed with the times. The hypotheses for this project are that music has become increasingly positive (*H1)* and music with more positive lyrics make it to the Billboard charts more often (*H2).* The results showed that all decades had a negative sentiment score. In addition, sentiment scores became increasingly negative after the 1970s decade. The lyrical sentiment behind the top Billboard songs seem to suggest that songs with negative lyrics are more welcomed than songs with positive lyrics.

**Introduction**

The present project looked at the sentiment of top Billboard songs over a span of 50 years. The purpose of this project was to decipher whether the lyrics from the most listened to music had an overall more positive or negative sentiment to them. The motivation for this topic stemmed from two things: world events and music being a universal art form. Firstly, I wanted to work with a dataset that spanned over a few decades to see if music changed with time. Naturally, different styles of music would be “in” for different time periods, but what I wanted to know was if the underlying meaning of the songs differed as well. This in conjunction with world events that were happening such as the Civil Rights movement, 9/11 terrorist attacks, increase in the number of school shootings, and natural disasters raised my curiosity as to whether musical lyrics have become more negative with the times.

The second motivation to work with this topic, specifically music, is that fact that music is the most universal music form. It seems as if everyone listens to music in one form or another. Because of this, I was intrigued by what kind of sentiment the most successful or popular music contained in their lyrics. In addition, music has interesting effects on an individual’s psychology. In a study done by Ziv, Chaim, and Itamar (2010), they found that listening to positive music led to participants feeling more hopeful and positive. Upon reading about this study, I wondered if music would have changed in a more positive way (contain more positive lyrics) as more negative events happened around the world. My curiosity was heightened by a different study by McCaffey (2008) whereby music was found to be an effective method of creating a healing environment in nursing homes. This study also notes that music is effective due to the fact that it is “a safe, inexpensive, and easy-to-use intervention” (McCaffrey 2008, pg 39). This study not only shows that music has positive psychological effect on people, but also notes the universality of music, quite fitting for my motivation to explore this topic. Bonny (1986) found similar results in her own study on music and healing.

In a separate study, the effect of music was also linked to emotional well-being in the long term (Campion & Levita, 2012). This may explain why it is such a popular form of art, but also interested me into asking the question of whether lyrics made a difference in effecting well-being. Would music with more positive lyrics garner more popularity? I initially questioned whether certain types of beats that are considered to be positive would also be more popular, but that is outside the scope of this project. Finally, the case for music having an effect on a person’s psychological well-being was further supported by a study done by Ventre (1994) who found that music could serve as a good component to healing childhood abuse.

From these background information, I have formalized three hypotheses: *H1:* Lyrics of most popular songs have become more positive with time to contradict or distract from the negative climate of world eventsand *H2:* More popular songs have more positive lyrics*.*

**Data**

The dataset used was taken from Kaggle.com. The dataset contained information on top billboard songs from 1965 to 2015. The data was put together by the user RakanNimer, who scraped the data from three distinct websites: metrolyrics.com, songlyrics.com, and lyricsmode.com. The dataset contained six variables: ‘Rank’, ‘Song’, ‘Artist’, ‘Year’, “Lyrics’, and ‘Source’. The variable ‘Rank’ is a column in this dataset that contains numeric values of the rank of the song in its respective year. For instance, if a song’s value in this column was ‘1’, it would mean that that song ranked number one in the year that the song was on the Billboard charts. The ‘Song’ variable contains character values of a song’s title. All of the values are in lowercase. The ‘Artist’ variable also holds values in the form of a character. These values represent the artist of the song and are all in lowercase as well. The ‘Year’ variable contains numerical values of the year that the song was on the Billboard charts. The range of year for this dataset is from 1965 to 2015. The ‘Lyrics’ variable holds lyrics of the song respective of the row in the form of character values. All values are in lowercase. Finally, the ‘Source’ variable contains numerical values that indicate which source the information of the song was pulled from, with ‘1’ being metrolyrics.com, ‘2’ being songlyrics.com, and ‘3’ being lyricsmode.com.

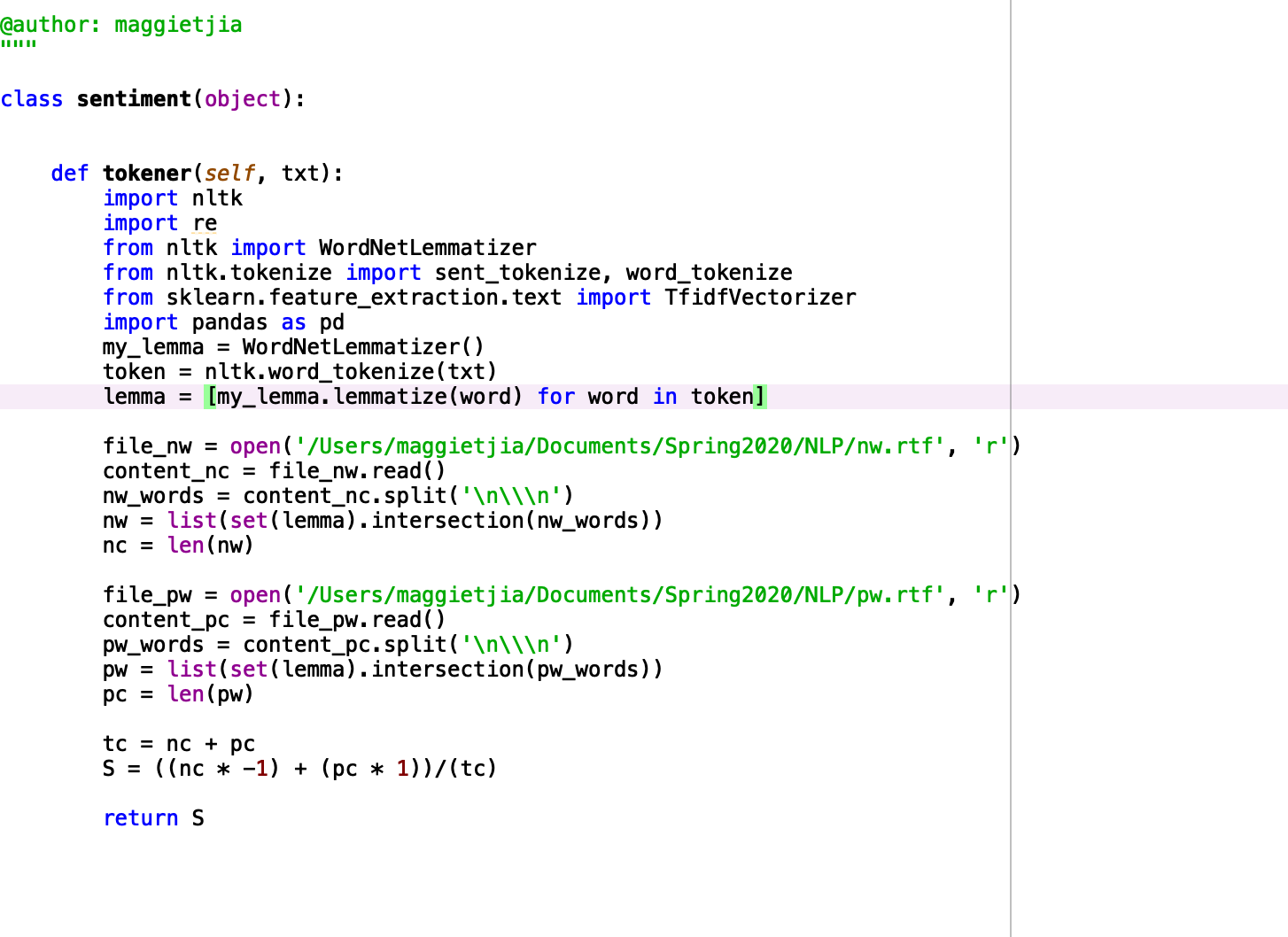
I also utilized the opinion lexicon by Hu and Liu (2004) as a measure of sentiment. This dictionary contains two lexicons that each contain a list of words. One lexicon represents positive sentiment, which contains words that are generally positive. The other lexicon contains negative words, representing negative sentiment.

**Methodology**

The present project was done completely in Spyder, an IDE for Python. I saved the downloaded datasets into a folder designated for this project. For this project, I created a sentiment analysis function (named ‘sentiment()’) that calculates a sentiment score for an object that is passed through the function. The sentiment analysis function requires the NLTK and RE packages. The basic form of the function tokenizes (via nltk.tokenize, specifically word\_tokenize) the text objects that are passed through. It then loads the negative sentiment file (represented by variable name ‘file\_nw’) from Hu and Liu (2004), reads the file (represented by ‘content\_nc’), splits the individual text elements into a separate list element (represented by ‘nw\_words’) and passes the tokenized object to see if there are any words that match the opinion lexicon (via .intersection, represented by ‘nw’). The result is then transformed into a list format and its length is calculated to see how many words match the negative lexicon (represented by ‘nc’). In other words, the length represents how many negative words the text object has. The same is done with the positive sentiment file (Hu and Liu, 2004; the variable names are the same as above with the exception being the ‘n’ being replaced by ‘p’ to represent the positive lexicon file).

The sentiment score is ultimately calculated by multiplying the ‘nc’ variable (number of negative words in the text object passed through the sentiment analysis function) by -1. This means that if a text object had 5 negative words found, it would have value of -5. Similarly, this is done with the ‘pc’ variable but multiplies by +1. The two values are added up and divided by a value called ‘tc’, which is the sum of ‘nc’ and ‘pc’, representing the total number of words in the text object that matches with both the positive and negative opinion lexicons by Hu and Liu (2004). The result of this calculation is set to the variable ‘S’, which gives the sentiment score of the text object that was passed through. This is to say, if a text object had nc = 5 and pc = 6, the value would be calculated as so: (-5 + 6) / 11 = 1/11 = 0.0909.

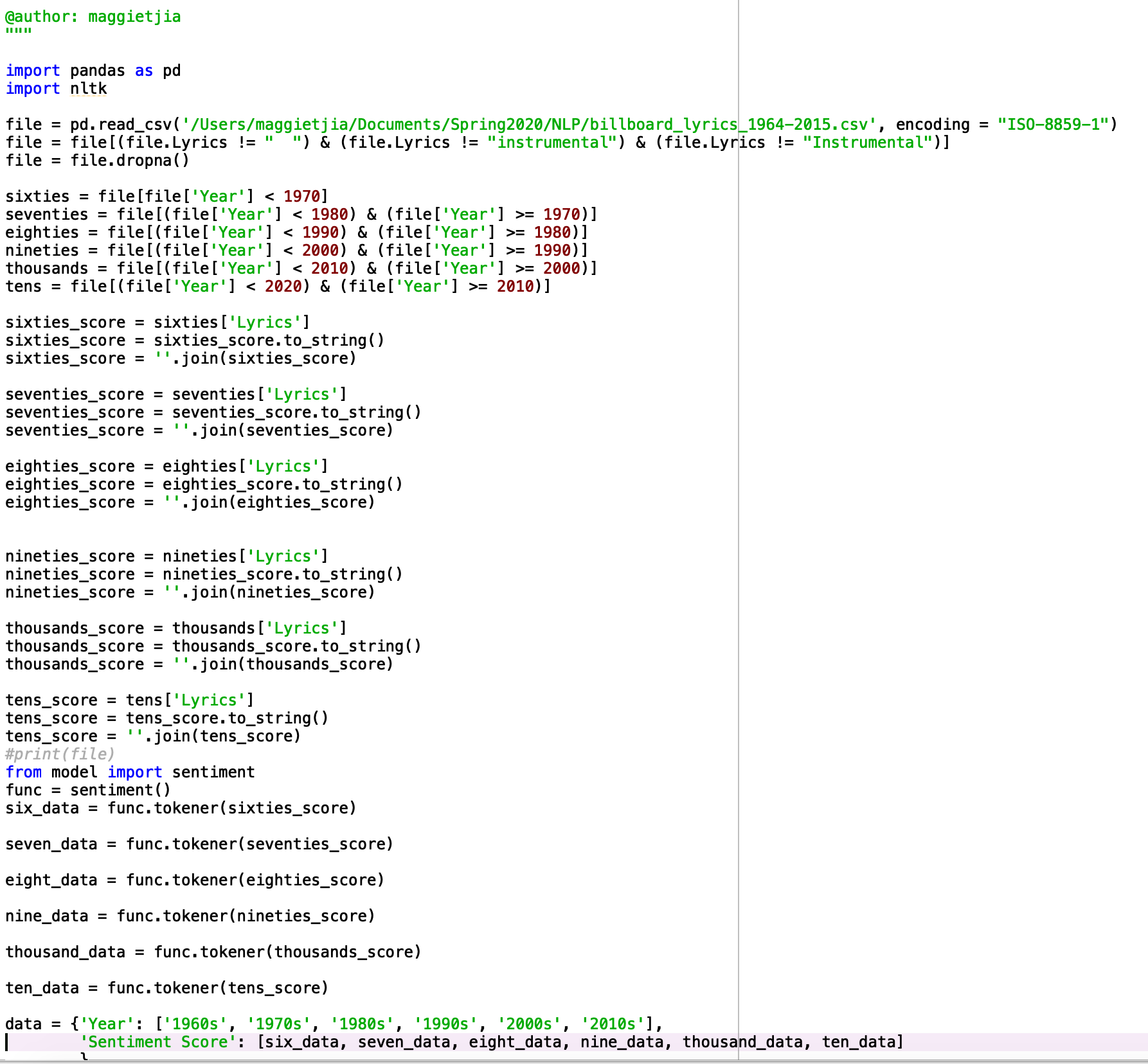
**Table 1**

*Sentiment analysis function with tokenization and lemmatization.*

The Billboard dataset was then loaded into a separate Spyder file. In this file, I also loaded in Pandas and NLTK. I read the dataset that was in .csv form via ‘pd.read\_csv’ and encoded it in ISO-8859-1 form. I then removed any songs that did not have any lyrics in the ‘Lyrics’ column or were instrumental (this meant the song had “Instrumental” in its ‘Lyrics’ column). I also removed any missing values from this dataset.

I split this dataset into decades. The 1960s was represented by songs that had ‘Year’ values of 1965 to 1969, the 1970s was represented by songs that had ‘Year’ values of 1970 to 1979, the 1980s was represented by songs that had ‘Year’ values of 1980 to 1989, the 1990s was represented by songs that had ‘Year’ values of 1990 to 1999, the 2000s was represented by songs that had ‘Year’ values of 2000 to 2009, and the 2010s was represented by songs that had ‘Year’ values of 2010 to 2015. For each decade, the ‘Lyrics’ column were transformed into strings and joined together to form one big string. This was to make it easier to pass into the sentiment() function, as all lyrics from that decade would be a single text object and represent the decade as one whole object.

**Table 2**



*Data wrangling of the Billboard dataset into respective decade data.*

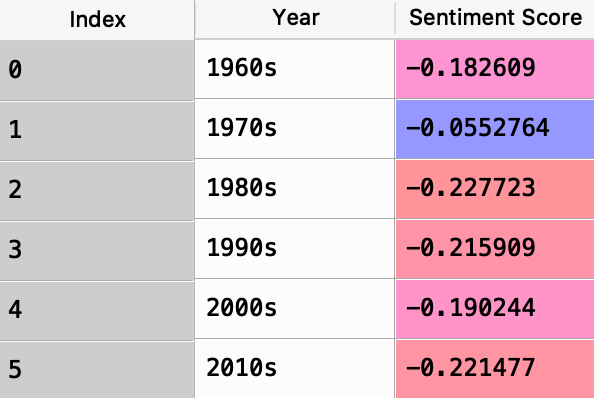
I then loaded in the sentiment() function into the Spyder file that contained the Billboard dataset. I ran the function for each decade separately and named them accordingly. After all the decades have been put through the sentiment() function, I gathered all the sentiment scores and put them in a dataframe in their own column. I created a separate column that displayed the decade that the sentiment score was for.

I repeated the process again, but added lemmatization to the sentiment() function. The process described above was repeated a third time with tokenization, lemmatization, as well as removal of stop words.

**Results**

The results from the first run through (sentiment scores from tokenizing the lyrics only) were negative. This means that for all decades that were analyzed in this project, their sentiment scores were all below zero. Notable results from the first set of results were that the decade of 1960s had a relatively low negative sentiment score, the decade of 1970s had the lowest negative sentiment score, and that sentiment scores were generally increasingly negative after the 1970s.

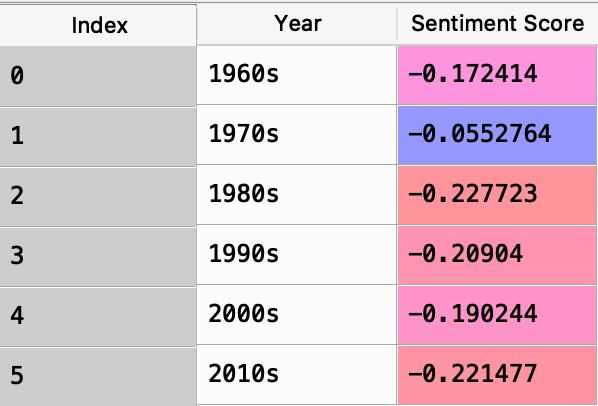
**Table 3**



*Sentiment scores for the first function that only gathered sentiment scores from tokenized text object.*

The results from the second that included lemmatizing the text objects were very similar to the first set of results. The minor differences in sentiment scores saw some decades have a more negative score (1960s to 1990s) and other decades have a more positive sentiment score (2000s and 2010s) when compared to the first function that only lemmatized the text objects. However, the pattern of the sentiment score changes from the first set of results remained the same in this second set of sentiment score results: The 1960s and 1970s saw a less negative set of sentiment scores while sentiment scores became more negative after 1970s.

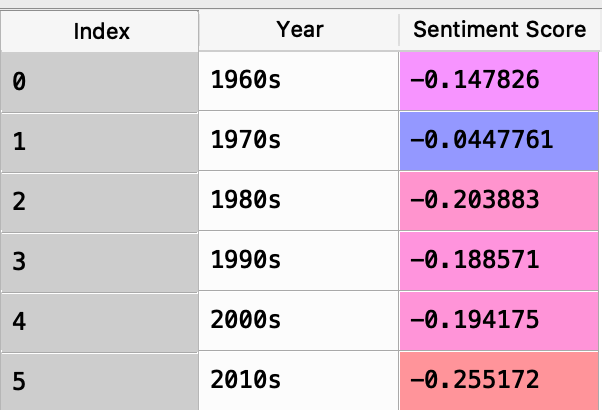
**Table 4**

****

*Sentiment scores for the second function that gathered sentiment scores from tokenized and lemmatized text object.*

The third set of sentiment scores are more different than the previous two sets of sentiment scores. In the third run through, stopwords were removed which may have contributed to a rather different sentiment score due to the fact that some of these stopwords may have been previously matched up with one of the opinion lexicons, resulting in a different sentiment score. However, the sentiment scores displayed a similar pattern like the previous two results: the 1960s and 1970s had the least negative sentiment scores whilst the sentiment scores beyond the 1970s saw a pattern of generally being more negative than the previous decades.

**Table 5**



*Sentiment scores for the third function that gathered sentiment scores from tokenized, lemmatized, and stopwords removed from text object.*

**Conclusion**

The current project aimed to look at whether lyrics to the most popular songs throughout five decades had changed in their sentiment. From the results, it is evident that there has been some sentiment change to top song lyrics. However, the results did not support my hypotheses, as all lyrics from the Billboard dataset had negative sentiment scores. This refutes my first hypothesis that lyrics for music became more positive as time went on (*H1*). A possible alternative explanation could be that musical artists sang about the negative world events rather than positive subjects to bring awareness to these situations. Alternatively, it could be that people simply enjoyed songs with more negative lyrics.

My second hypothesis that more popular songs have more positive lyrics was also refuted by all decades having a negative sentiment score (*H2)*. Like before, it could be that songs with a more negative kick to their lyrics are better received than songs that have a more positive meaning to them. It could also be due to the fact that the 1960s and 2010s were not fairly represented, as they only had 5 years that represented the whole decade. The somewhat incomplete dataset could have misrepresented what the sentiment of these decades really were.

The overall conclusion of this project is that the most popular music, at least the ones that make it onto Billboard charts, are generally negative in their lyrical nature. A pattern that stands out on top of that is the increasingly negative sentiment these lyrics have after the 1970s, suggesting that these types of songs were more well received than songs with more positive lyrics.

**References**

Bonny, H. L. (1986). Music and healing. *Music Therapy, 6(1),* 3 – 12.

Campion, M. & Levita, L. (2012). Enhancing positive affect and divergent thinking

abilities: Play some music and dance. *The Journal of Positive Psychology, 9(2)*,

137 – 145.

Hu, M. & Liu, B. (2004). Mining and summarizing customer reviews.*Proceedings of*

*the tenth ACM SIGKDD international conference on Knowledge discovery and*

*data mining.* 169 – 177.

McCaffrey, R. (2008). Music listening: Its effects in creating a healing environment.

*Journal of Psychosocial Nursing & Mental Health Sercives, 46(10),* 39 – 45.

Ventre, M. E. (1994). Healing the wounds of childhood abuse: A guided imagery and

music case study. *Music Therapy Perspectives, 12,* 98 – 103.

Ziv, N., Chaim, A. B., & Itamar, O. (2010). The effect of positive music and dispositional

hope on state hope and affect. *Psychology of Music, 39,*  3 – 17.