Master Project for
Major Social Data Science

Success in music Production: Analysis of Hot 100 Songs Over Time

By: Sayedmahdi Raghib

Sayedmahdi.raghib@rwth-aachen.de

Advisor : Prof. Jürgen Lerner
Date: 04.07.2023

RWTH Aachen University
Chair for Applied Computational Social Sciences

Contents

- Objective
- Approach
- Results
- Reflection
- Limitations
- Future Work

Objective

Research Questions

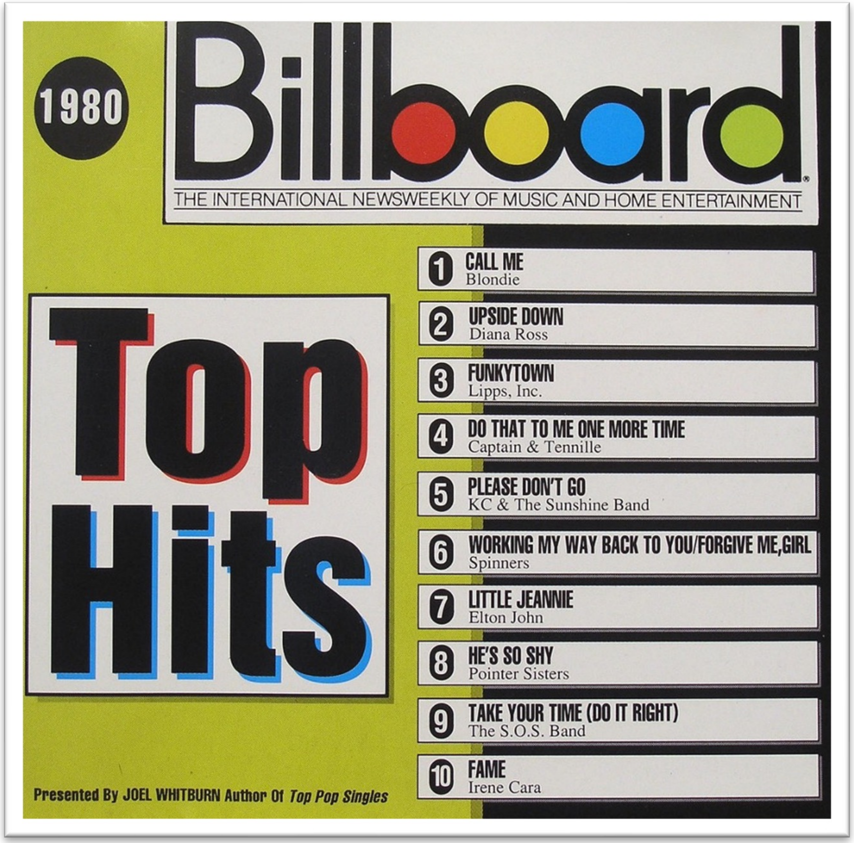
Objective

The goal of this analysis is to provide valuable insights to music industry players on what makes a song successful:

- Analyze the **successful** songs that made it to the Weekly Hot 100 singles chart by Billboard with other **unsuccessful** songs that are not in the chart to examine their **characteristics** such as genre, tempo and danceability.
- **Sentiment analysis** on both successful and unsuccessful songs
- **Top keyword analysis** on lyrics of the songs

Billboard Hot 100

- Billboard Hot 100 is the music industry standard record chart in the United States for songs, published weekly by Billboard magazine.
- Chart rankings are based on sales (physical and digital), radio play, and online streaming in the United States.


1980
Billboard
THE INTERNATIONAL NEWSWEEKLY OF MUSIC AND HOME ENTERTAINMENT
Top Hits
Presented By JOEL WHITBURN Author Of Top Pop Singles

1 CALL ME
Blondie

2 UPSIDE DOWN
Diana Ross

3 FUNKYTOWN
Lipps, Inc.

4 DO THAT TO ME ONE MORE TIME
Captain & Tennille

5 PLEASE DON'T GO
KC & The Sunshine Band

6 WORKING MY WAY BACK TO YOU/FORGIVE ME,GIRL
Spinners

7 LITTLE JEANNIE
Elton John

8 HE'S SO SHY
Pointer Sisters

9 TAKE YOUR TIME (DO IT RIGHT)
The S.O.S. Band

10 FAME
Irene Cara

Research Questions

1. What is a pattern of the longevity of hot songs? (Some hit songs appear in the list once and some more)
2. What patterns of repeated success among the artists? (Unique hit songs released by artists.)
3. What is the temporal trend of the audio features?
    - Do tracks become more danceable or more relaxed over the past years based on the audio attribute levels?
    - Is there a seasonal mood change?

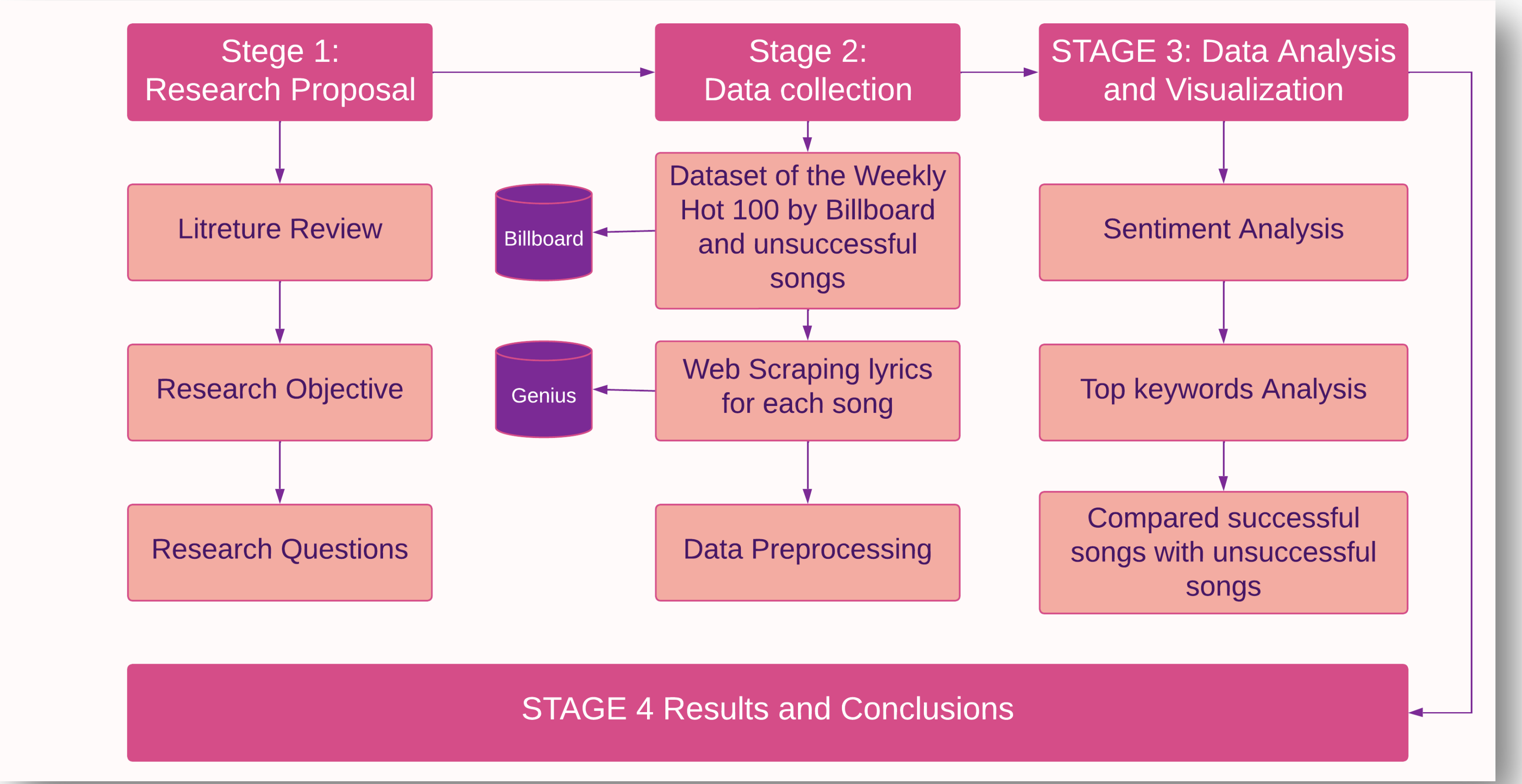
Research Questions

5. How has the sentiment of lyrics in popular music changed over time?
6. What are the top keywords that have been consistently present in popular song lyrics over time?
7. Is there any correlation between the sentiment of the song's lyrics and its chart performance? Do More positive or negative songs tend to be more successful?
8. How has the vocabulary used in popular music changed over time? Are certain words more common in certain eras?

Approach

How I have tackled the task

Research Framework


Stege 1: Research Proposal
Stage 2: Data collection
STAGE 3: Data Analysis and Visualization
Litreture Review
Billboard
Dataset of the Weekly Hot 100 by Billboard and unsuccessful songs
Sentiment Analysis
Research Objective
Genius
Web Scraping lyrics for each song
Top keywords Analysis
Research Questions
Data Preprocessing
Compared successful songs with unsuccessful songs
STAGE 4 Results and Conclusions

Data Description

Successful songs dataset provided us with information every weekly Hot 100 singles chart from Billboard.com from 1958 to 2019. (data.world):

- Total number of successful songs is 21741.

Unsuccessful songs dataset provided us with any songs that have never appeared in the weekly Hot 100 singles chart from Billboard.com. (Kaggle.com)

- Total number of unsuccessful songs is 23671.

Data Description

Audio Features

- **acousticness -** A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
- **danceability -** Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.
- **energy -** Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale.

Data Description

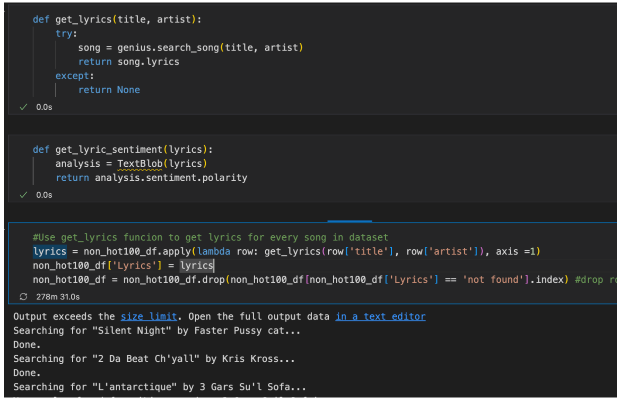
Audio Features

- **instrumentalness** - Predicts whether a track contains no vocals. “Ooh” and “aah” sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly “vocal”. The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content.
- **loudness** - The overall loudness of a track in decibels (dB). Values typical range between -60 and 0 db.
- **valence** - A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric).

Approach

Lyrics Scraping

- I scraped the lyrics for the songs from the website genius.com using a python API client called LyricsGenius.



```
def get_lyrics(title, artist):
    try:
        song = genius.search_song(title, artist)
        return song.lyrics
    except:
        return None
```

✓ 0.0s

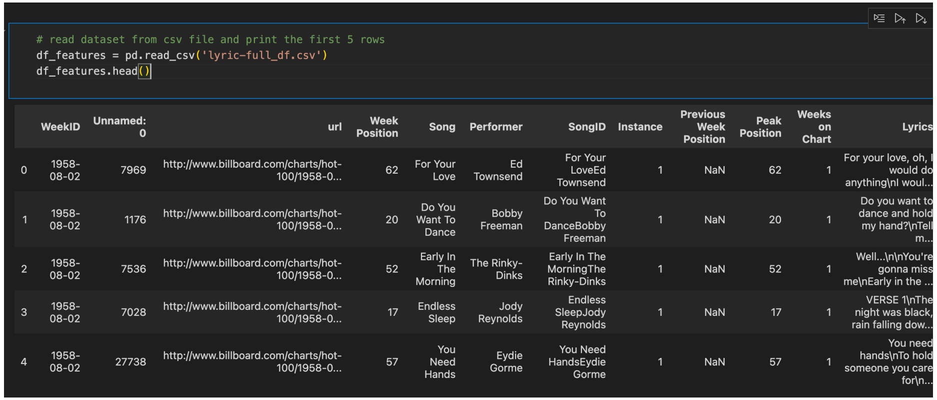
```
def get_lyric_sentiment(lyrics):
    analysis = TextBlob(lyrics)
    return analysis.sentiment.polarity
```

✓ 0.0s

```
#Use get_lyrics funcion to get lyrics for every song in dataset
lyrics = non_hot100_df.apply(lambda row: get_lyrics(row['title'], row['artist']), axis =1)
non_hot100_df['Lyrics'] = lyrics
non_hot100_df = non_hot100_df.drop(non_hot100_df[non_hot100_df['Lyrics'] == 'not found'].index) #drop r
```

278m 31.0s

```
Output exceeds the size limit. Open the full output data in a text editor
Searching for "Silent Night" by Faster Pussy cat...
Done.
Searching for "2 Da Beat Ch'yall" by Kris Kross...
Done.
Searching for "L'antarctique" by 3 Gars Su'l Sofa...
```



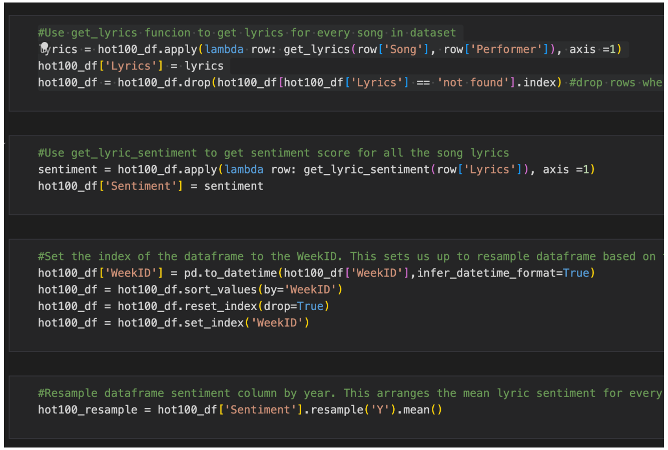
```
# read dataset from csv file and print the first 5 rows
df_features = pd.read_csv('lyric-full_df.csv')
df_features.head()
```

	WeekID	Unnamed: 0	url	Week Position	Song	Performer	SongID	Instance	Previous Week Position	Peak Position	Weeks on Chart	Lyrics
0	1958-08-02	7969	http://www.billboard.com/charts/hot-100/1958-0...	62	For Your Love	Ed Townsend	For Your LoveEd Townsend	1	NaN	62	1	For your love, oh, I would do anything\nI woul...
1	1958-08-02	1176	http://www.billboard.com/charts/hot-100/1958-0...	20	Do You Want To Dance	Bobby Freeman	Do You Want To DanceBobby Freeman	1	NaN	20	1	Do you want to dance and hold my hand?\nTell m...
2	1958-08-02	7536	http://www.billboard.com/charts/hot-100/1958-0...	52	Early In The Morning	The Rinky-Dinks	Early In The MorningThe Rinky-Dinks	1	NaN	52	1	Well...\n\nYou're gonna miss me\nEarly in the ...
3	1958-08-02	7028	http://www.billboard.com/charts/hot-100/1958-0...	17	Endless Sleep	Jody Reynolds	Endless SleepJody Reynolds	1	NaN	17	1	VERSE 1\nThe night was black, rain falling dow...
4	1958-08-02	27738	http://www.billboard.com/charts/hot-100/1958-0...	57	You Need Hands	Eydie Gorme	You Need HandsEydie Gorme	1	NaN	57	1	You need hands\nTo hold someone you care for\n...

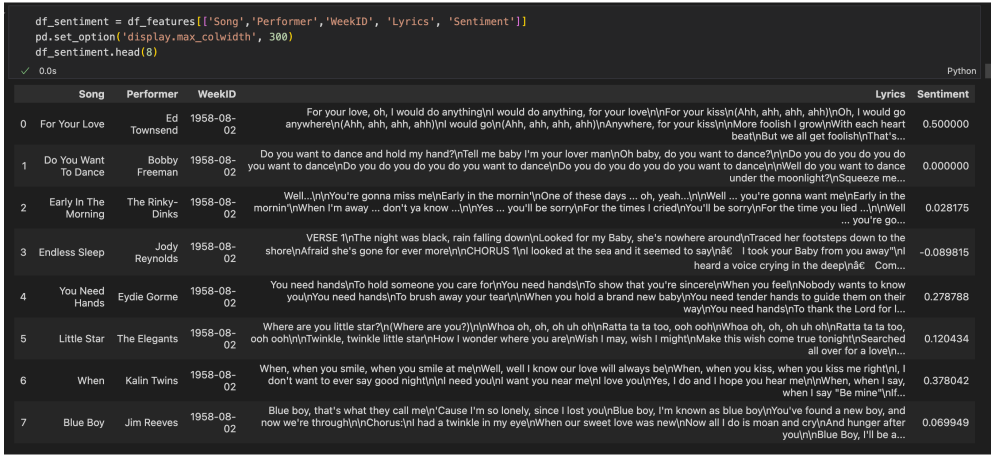
Approach

Sentiment Analysis

- Performed sentiment analysis on the lyrics used TextBlob library.



```
#Use get_lyrics funcion to get lyrics for every song in dataset
lyrics = hot100_df.apply(lambda row: get_lyrics(row['Song'], row['Performer']), axis =1)
hot100_df['Lyrics'] = lyrics
hot100_df = hot100_df.drop(hot100_df[hot100_df['Lyrics'] == 'not found'].index) #drop rows whe

#Use get_lyric_sentiment to get sentiment score for all the song lyrics
sentiment = hot100_df.apply(lambda row: get_lyric_sentiment(row['Lyrics']), axis =1)
hot100_df['Sentiment'] = sentiment

#Set the index of the dataframe to the WeekID. This sets us up to resample dataframe based on
hot100_df['WeekID'] = pd.to_datetime(hot100_df['WeekID'],infer_datetime_format=True)
hot100_df = hot100_df.sort_values(by='WeekID')
hot100_df = hot100_df.reset_index(drop=True)
hot100_df = hot100_df.set_index('WeekID')

#Resample dataframe sentiment column by year. This arranges the mean lyric sentiment for every
hot100_resample = hot100_df['Sentiment'].resample('Y').mean()
```



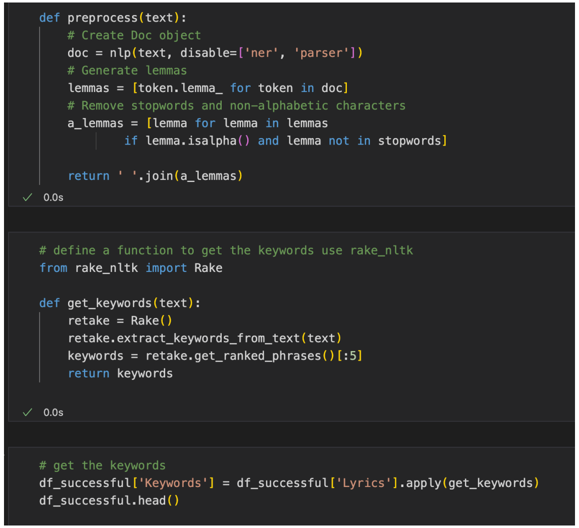
```
df_sentiment = df_features[['Song','Performer','WeekID', 'Lyrics', 'Sentiment']]
pd.set_option('display.max_colwidth', 300)
df_sentiment.head(8)
```

	Song	Performer	WeekID	Lyrics	Sentiment
0	For Your Love	Ed Townsend	1958-08-02	For your love, oh, I would do anything\nI would do anything, for your love\n\nFor your kiss\n(Ahh, ahh, ahh, ahh)\nOh, I would go anywhere\n(Ahh, ahh, ahh, ahh)\nI would go\n(Ahh, ahh, ahh, ahh)\nAnywhere, for your kiss\n\nMore foolish I grow\nWith each heart beat\nBut we all get foolish\nThat's...	0.500000
1	Do You Want To Dance	Bobby Freeman	1958-08-02	Do you want to dance and hold my hand?\nTell me baby I'm your lover man\nOh baby, do you want to dance?\n\nDo you do you do you do you want to dance\nDo you do you do you do you want to dance\nDo you do you do you do you want to dance\n\nWell do you want to dance under the moonlight?\nSqueeze me...	0.000000
2	Early In The Morning	The Rinky-Dinks	1958-08-02	Well...\n\nYou're gonna miss me\nEarly in the mornin'\nOne of these days ... oh, yeah...\n\nWell ... you're gonna want me\nEarly in the mornin'\nWhen I'm away ... don't ya know ...\n\nYes ... you'll be sorry\nFor the times I cried\nYou'll be sorry\nFor the time you lied ...\n\nWell ... you're go...	0.028175
3	Endless Sleep	Jody Reynolds	1958-08-02	VERSE 1\nThe night was black, rain falling down\nLooked for my Baby, she's nowhere around\nTraced her footsteps down to the shore\nAfraid she's gone for ever more\n\nCHORUS 1\nI looked at the sea and it seemed to say\nâ€ I took your Baby from you away"\nI heard a voice crying in the deep\nâ€ Com...	-0.089815
4	You Need Hands	Eydie Gorme	1958-08-02	You need hands\nTo hold someone you care for\nYou need hands\nTo show that you're sincere\nWhen you feel\nNobody wants to know you\nYou need hands\nTo brush away your tear\n\nWhen you hold a brand new baby\nYou need tender hands to guide them on their way\nYou need hands\nTo thank the Lord for l...	0.278788
5	Little Star	The Elegants	1958-08-02	Where are you little star?\n(Where are you?)\n\nWhoa oh, oh, oh uh oh\nRatta ta ta too, ooh ooh\nWhoa oh, oh, oh uh oh\nRatta ta ta too, ooh ooh\n\nTwinkle, twinkle little star\nHow I wonder where you are\nWish I may, wish I might\nMake this wish come true tonight\nSearched all over for a love\n...	0.120434
6	When	Kalin Twins	1958-08-02	When, when you smile, when you smile at me\nWell, well I know our love will always be\nWhen, when you kiss, when you kiss me right\nI, I don't want to ever say good night\n\nI need you\nI want you near me\nI love you\nYes, I do and I hope you hear me\n\nWhen, when I say, when I say "Be mine"\nIf...	0.378042
7	Blue Boy	Jim Reeves	1958-08-02	Blue boy, that's what they call me\n'Cause I'm so lonely, since I lost you\nBlue boy, I'm known as blue boy\nYou've found a new boy, and now we're through\n\nChorus:\nI had a twinkle in my eye\nWhen our sweet love was new\nNow all I do is moan and cry\nAnd hunger after you\n\nBlue Boy, I'll be a...	0.069949

Approach

Keyword Extraction

- Extraction the keywords from lyrics of each song by combining the Rapid Automatic Keyword Extraction (RAKE) algorithm with the NLTK toolkit, known as rake-nltk.



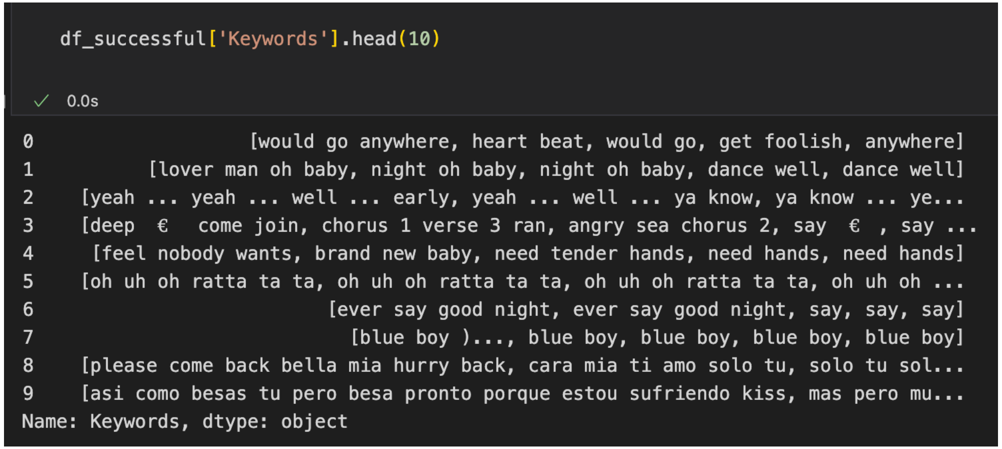
```python
def preprocess(text):
    # Create Doc object
    doc = nlp(text, disable=['ner', 'parser'])
    # Generate lemmas
    lemmas = [token.lemma_ for token in doc]
    # Remove stopwords and non-alphabetic characters
    a_lemmas = [lemma for lemma in lemmas
            if lemma.isalpha() and lemma not in stopwords]

    return ' '.join(a_lemmas)
```

✓ 0.0s

```python
# define a function to get the keywords use rake_nltk
from rake_nltk import Rake

def get_keywords(text):
    retake = Rake()
    retake.extract_keywords_from_text(text)
    keywords = retake.get_ranked_phrases()[:5]
    return keywords
```

✓ 0.0s

```python
# get the keywords
df_successful['Keywords'] = df_successful['Lyrics'].apply(get_keywords)
df_successful.head()
```



```python
df_successful['Keywords'].head(10)
```

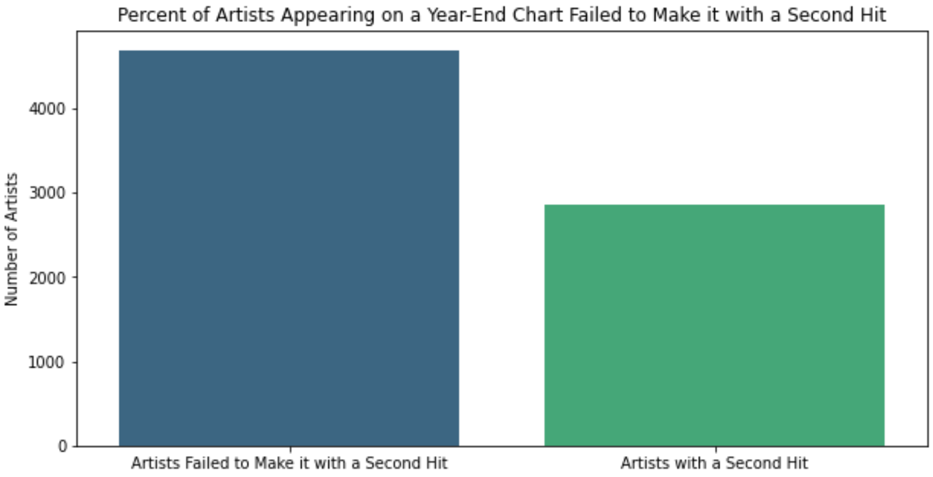
✓ 0.0s

```
0                   [would go anywhere, heart beat, would go, get foolish, anywhere]
1              [lover man oh baby, night oh baby, night oh baby, dance well, dance well]
2    [yeah ... yeah ... well ... early, yeah ... well ... ya know, ya know ... ye...
3    [deep  €   come join, chorus 1 verse 3 ran, angry sea chorus 2, say  €  , say ...
4     [feel nobody wants, brand new baby, need tender hands, need hands, need hands]
5    [oh uh oh ratta ta ta, oh uh oh ratta ta ta, oh uh oh ratta ta ta, oh uh oh ...
6                          [ever say good night, ever say good night, say, say, say]
7                            [blue boy )..., blue boy, blue boy, blue boy, blue boy]
8    [please come back bella mia hurry back, cara mia ti amo solo tu, solo tu sol...
9    [asi como besas tu pero besa pronto porque estou sufriendo kiss, mas pero mu...
Name: Keywords, dtype: object
```

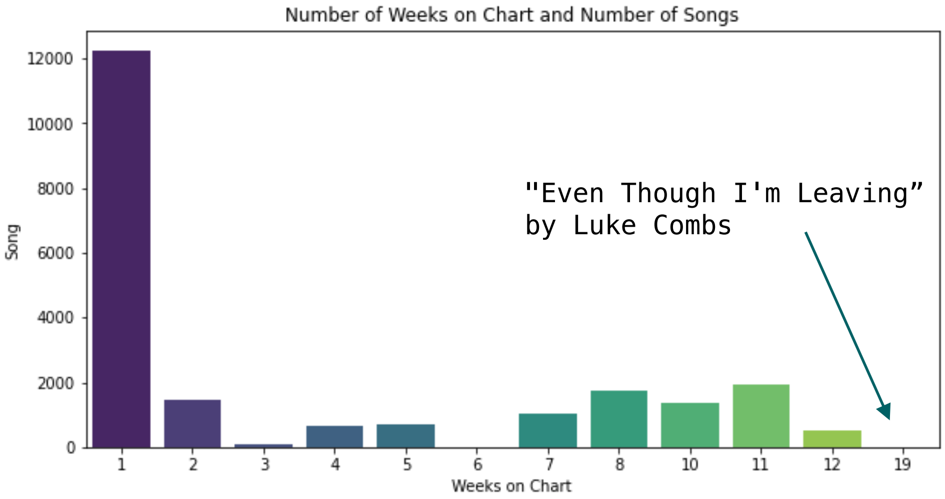
Results

Results

- What is a pattern of the longevity of hot songs?



Percent of Artists Appearing on a Year-End Chart Failed to Make it with a Second Hit

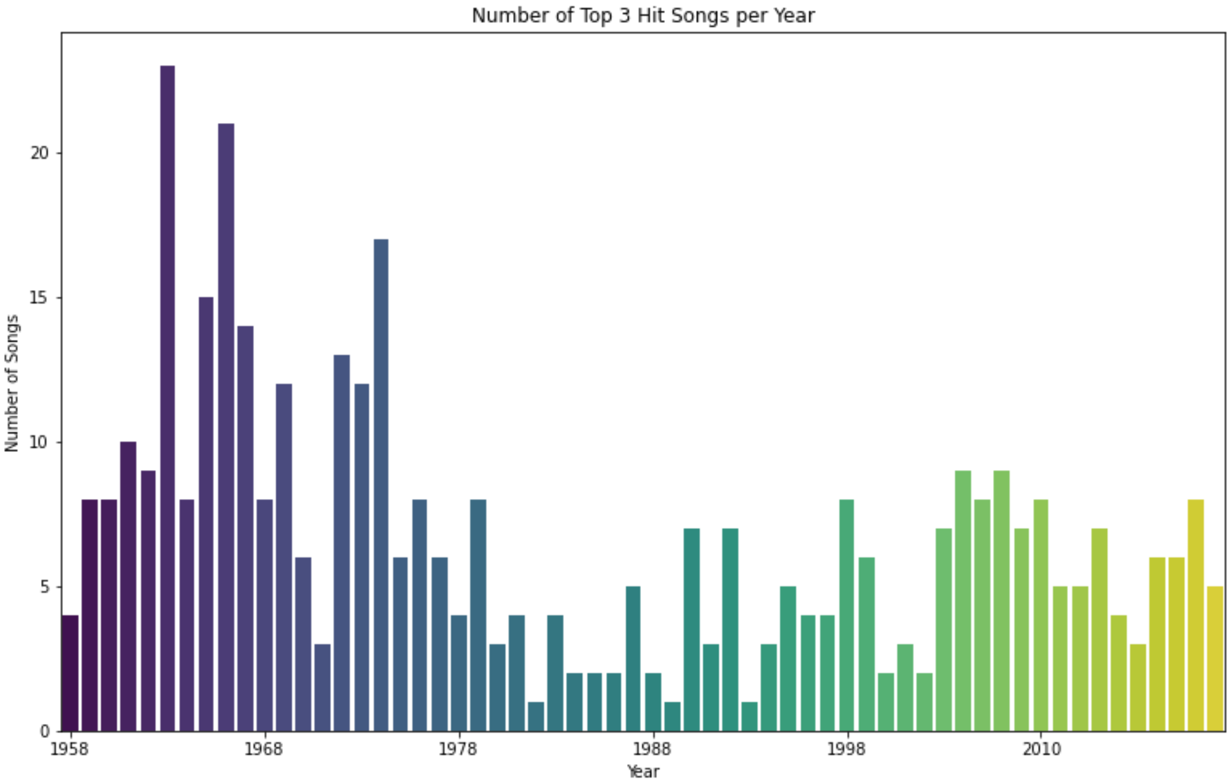


Number of Weeks on Chart and Number of Songs

"Even Though I'm Leaving”
by Luke Combs

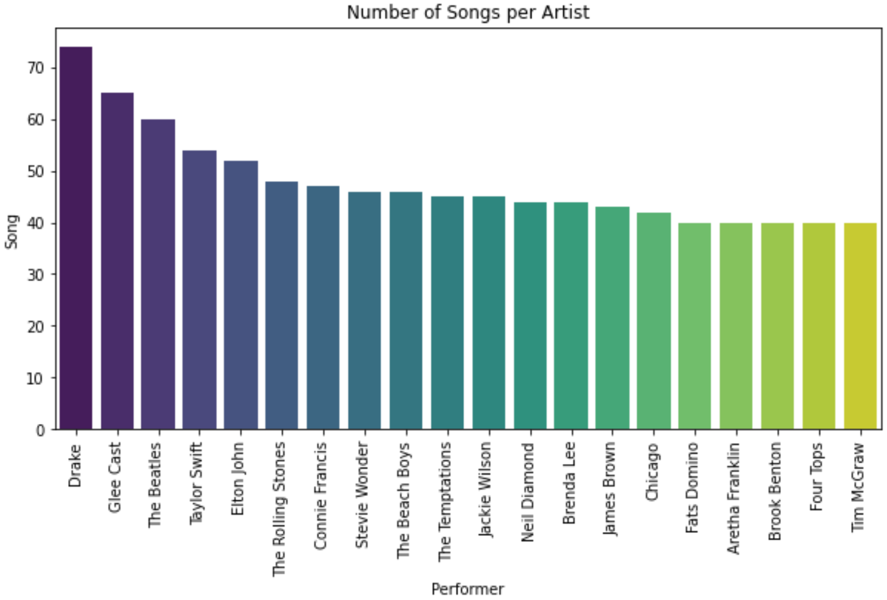
Results

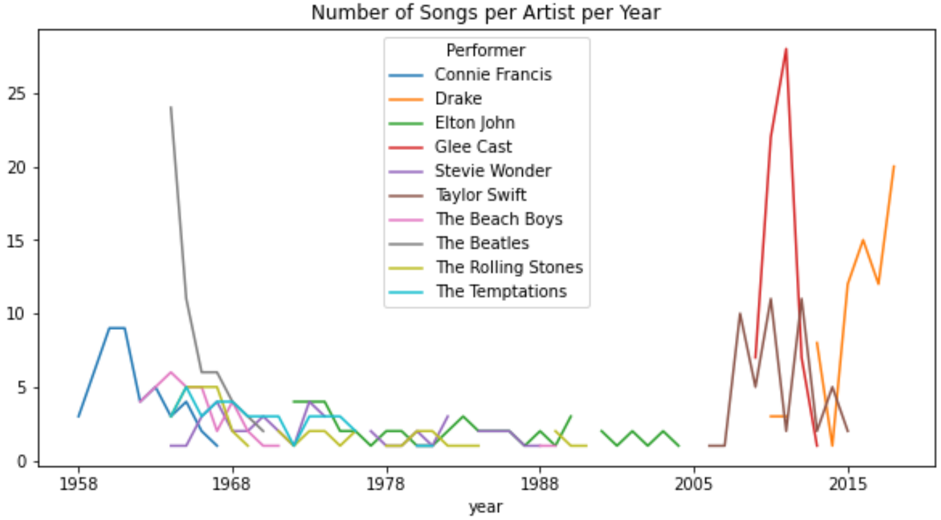
- What is the number of top hits on Billboard over time ?



Results

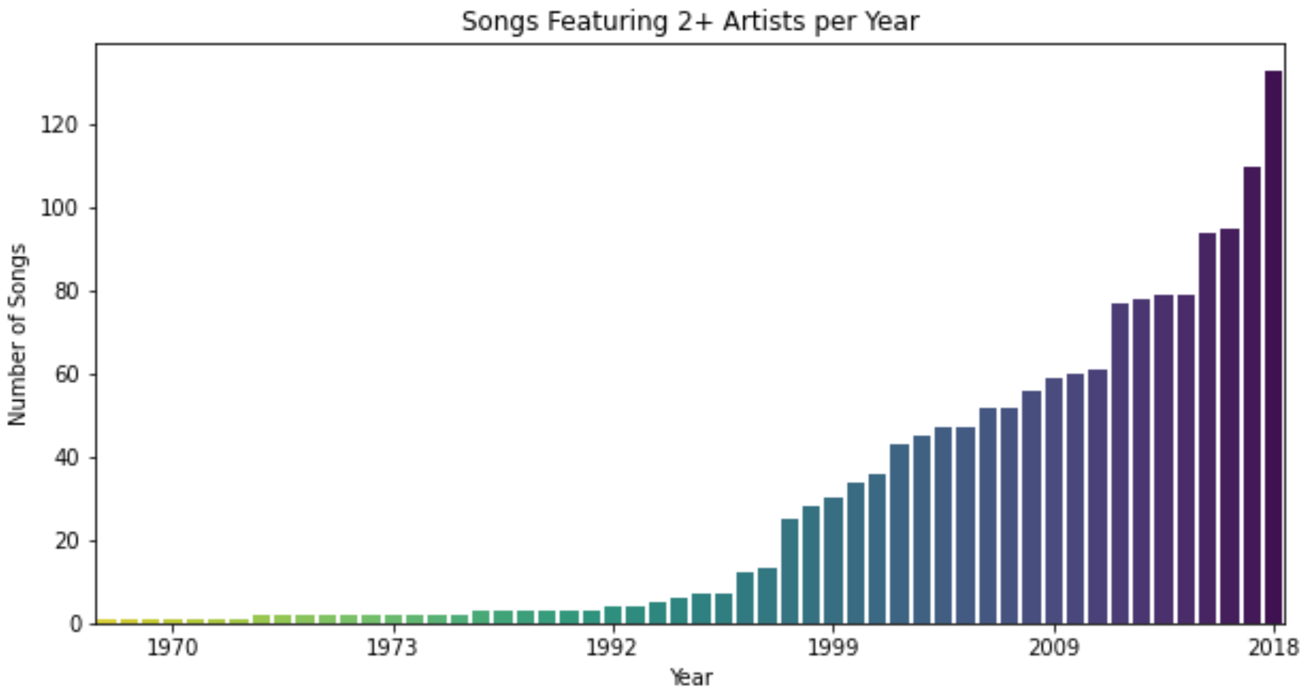
- What patterns of repeated success among the artists?





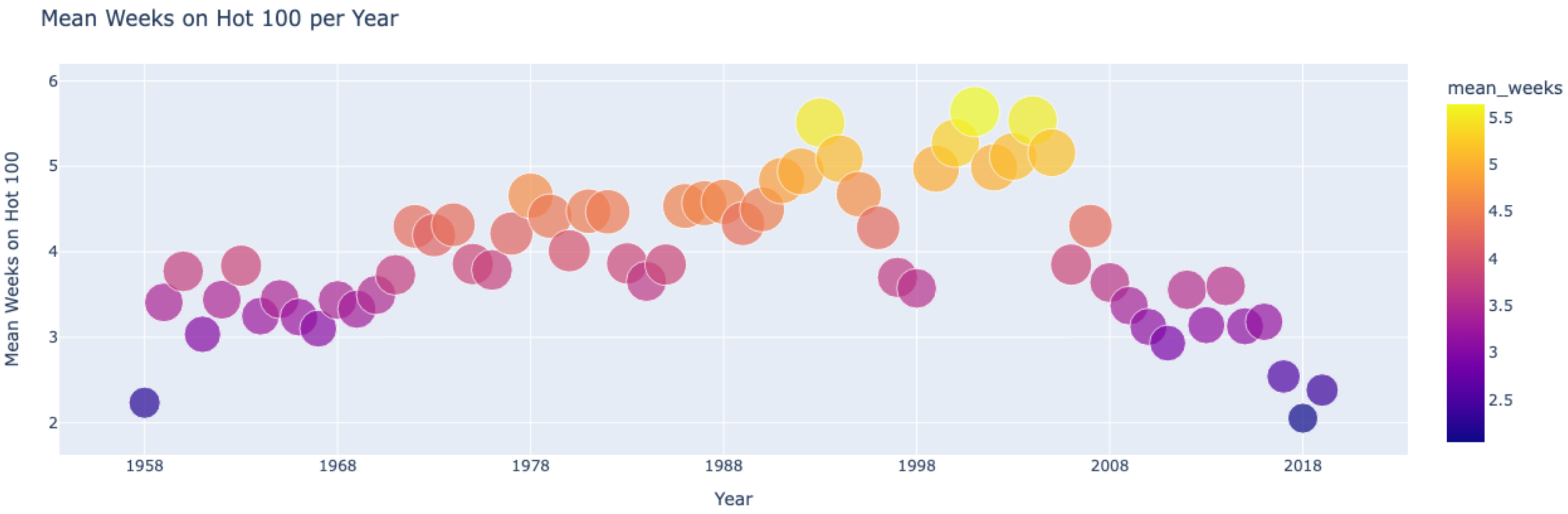
Results

- What patterns of repeated success among the artists?



Results

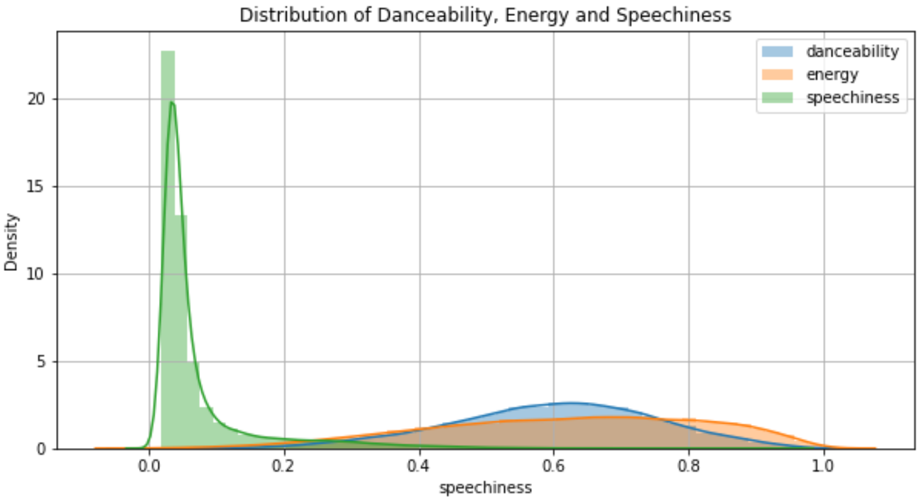
- What patterns of repeated success among the artists?



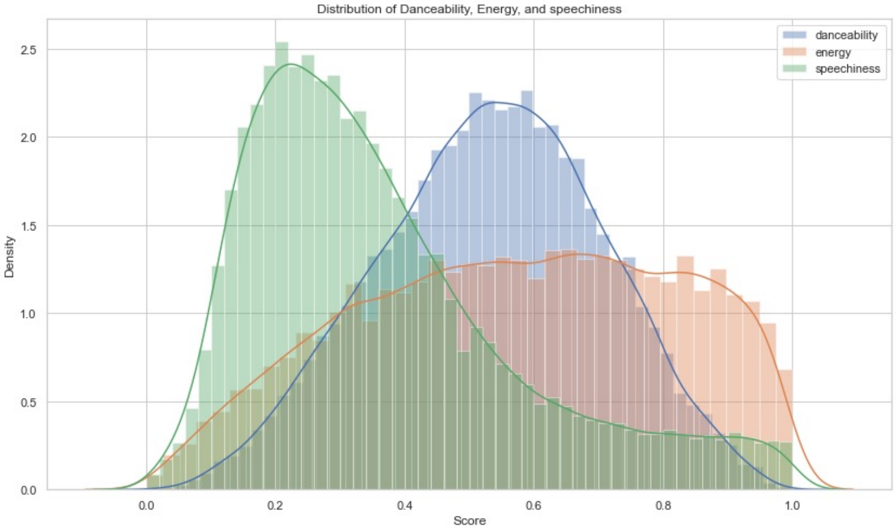
Results

- What is the temporal trend of the audio features?

Successful Songs:



Unsuccessful Songs:

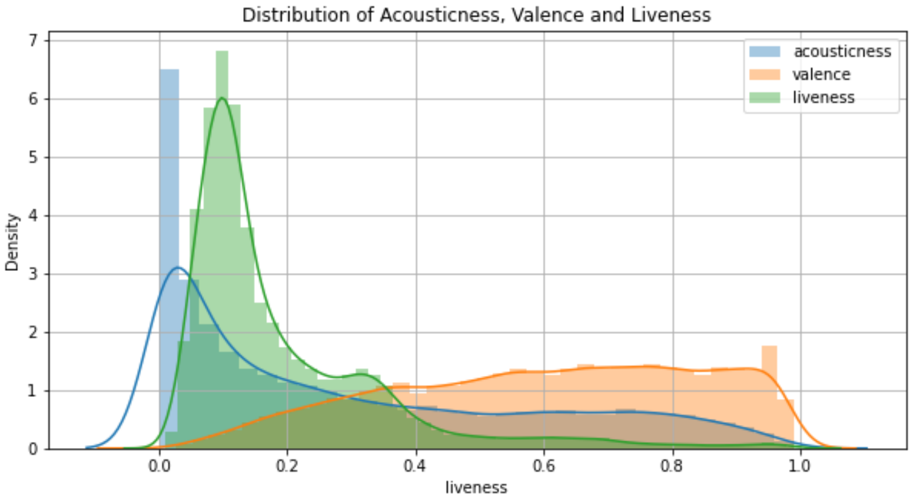


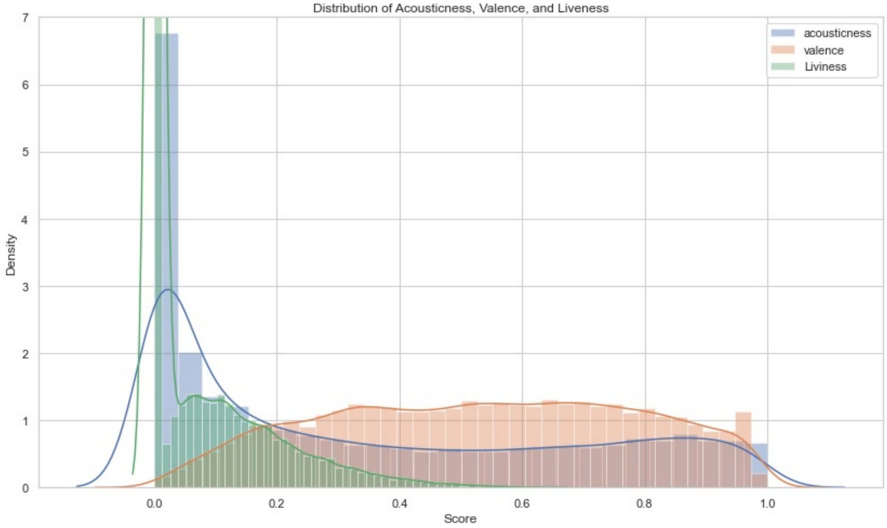
Results

- What is the temporal trend of the audio features?

Successful Songs:

Unsuccessful Songs:

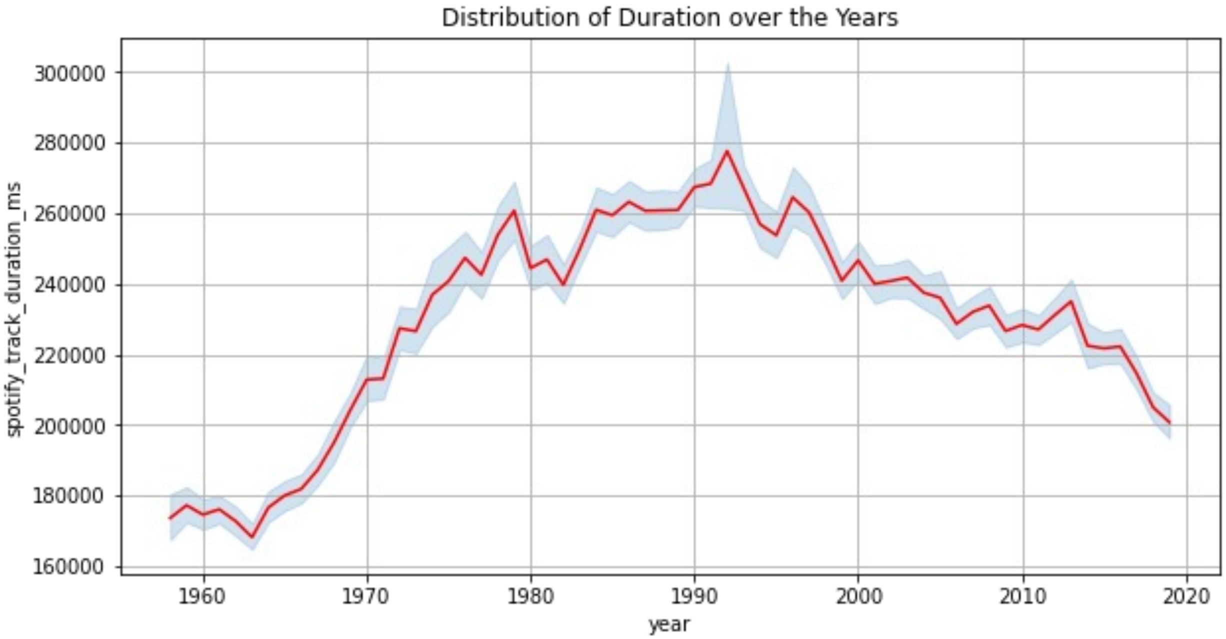




Results

- What is the temporal trend of the audio features?

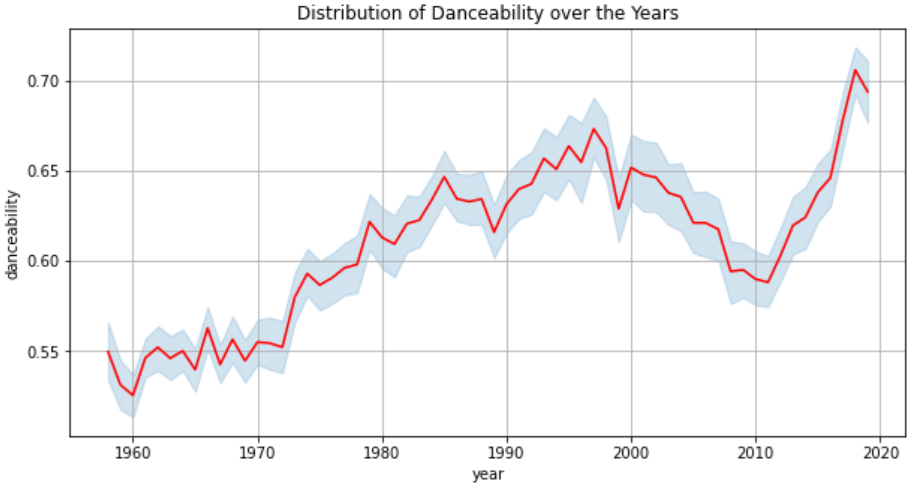
Successful Songs:



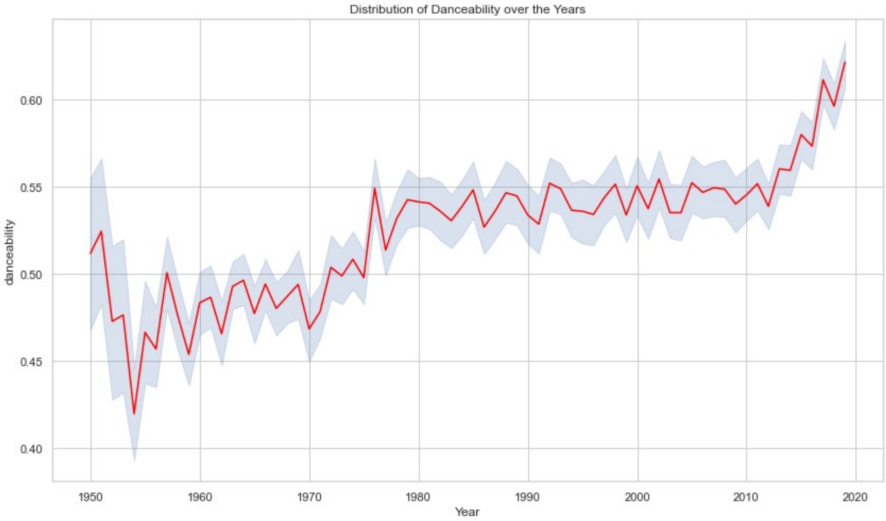
Results

- Do tracks become more danceable ?

Successful Songs:



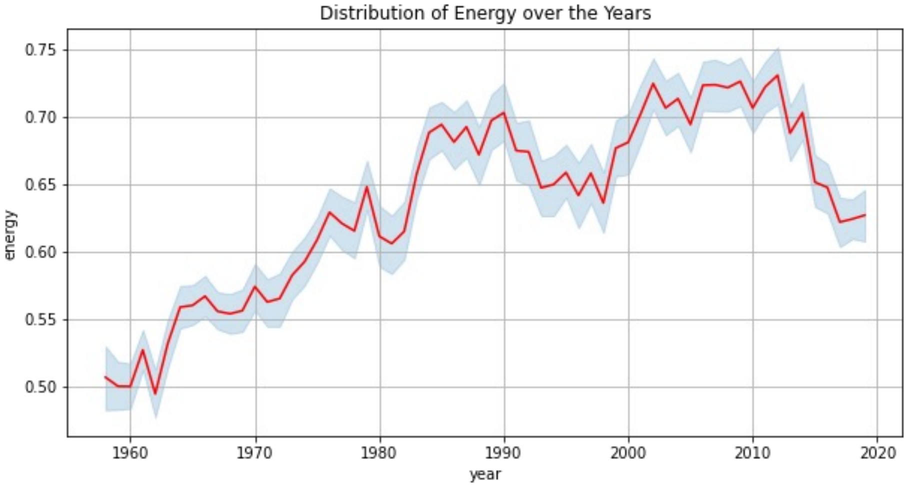
Unsuccessful Songs:



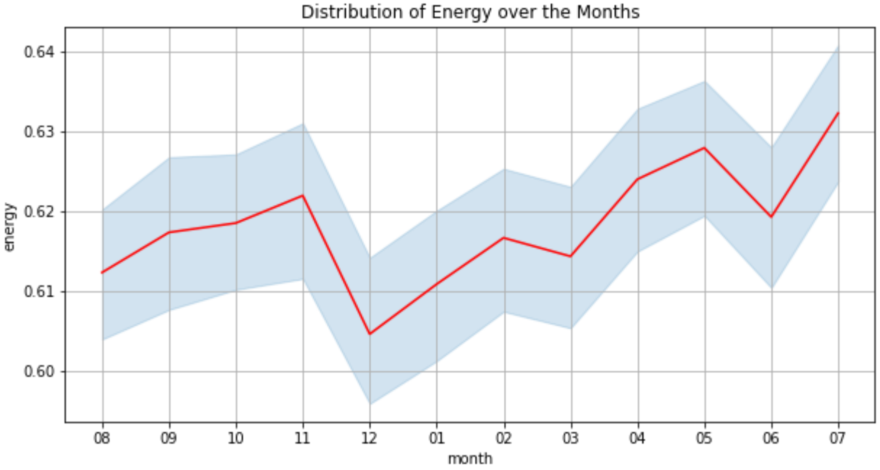
Results

- Is there a seasonal mood change in the successful songs?

Over the years:



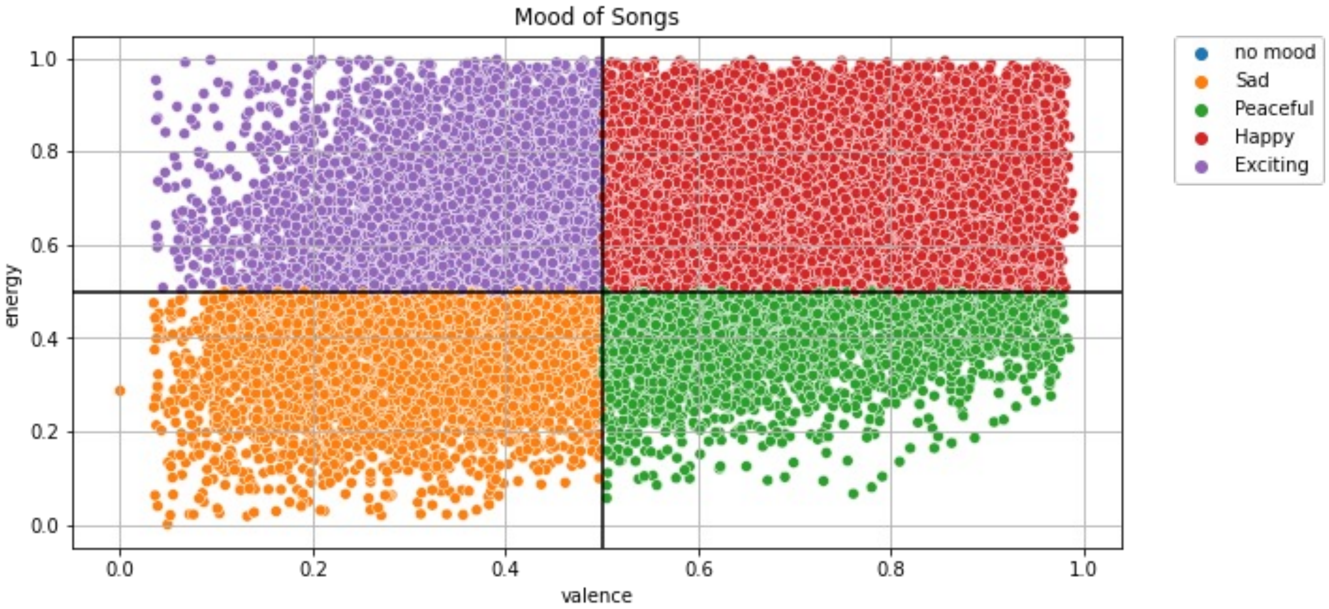
Over the months:



Results

- Mood of successful songs:

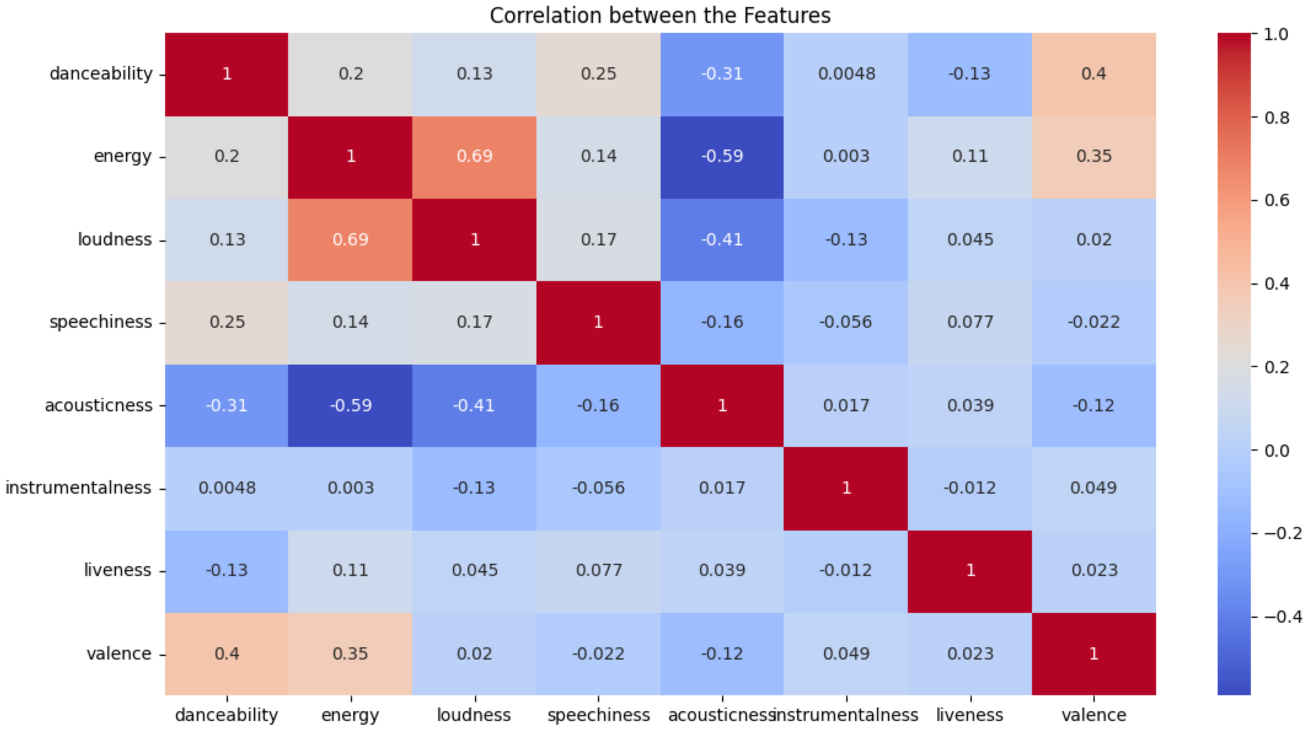
- happy: valence > 0.5, energy > 0.5
- excited: valence <= 0.5, energy > 0.5
- sad: valence <= 0.5, energy <= 0.5
- peaceful: valence > 0.5, energy <= 0.5.



Results

- What is the correlation between the audio features?

Successful Songs:



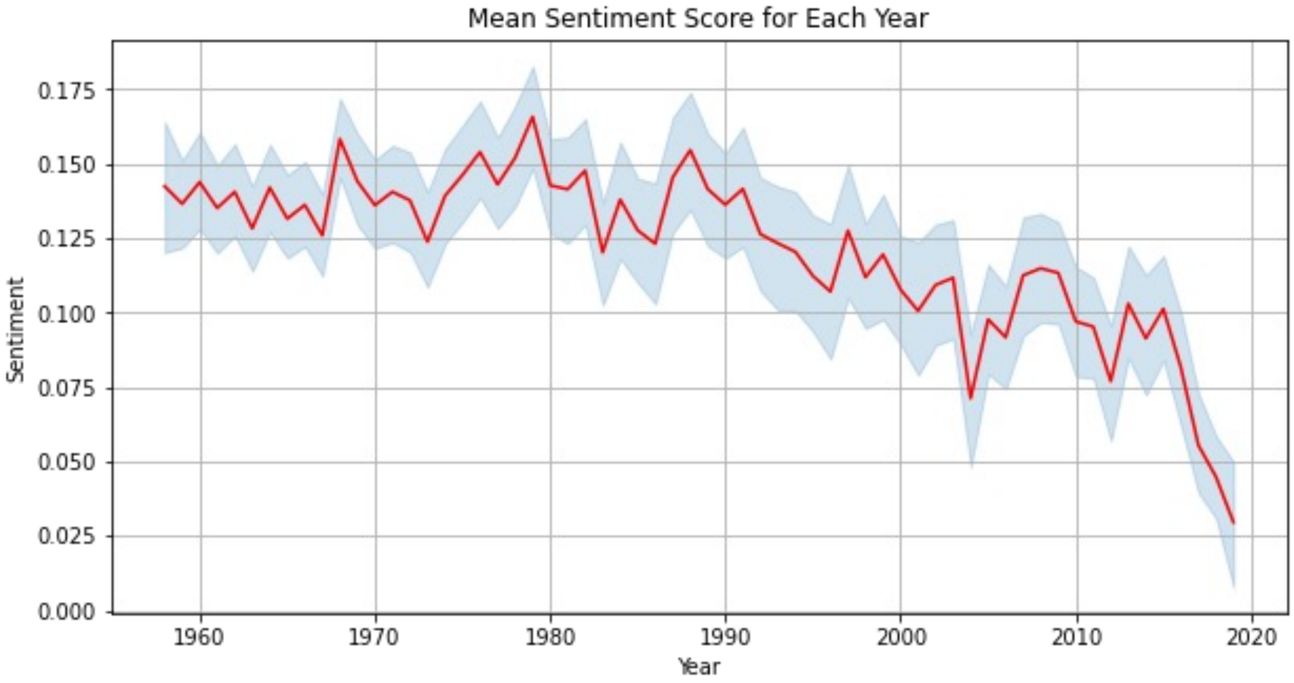
Correlation between the Features

	danceability	energy	loudness	speechiness	acousticness	instrumentalness	liveness	valence
danceability	1	0.2	0.13	0.25	-0.31	0.0048	-0.13	0.4
energy	0.2	1	0.69	0.14	-0.59	0.003	0.11	0.35
loudness	0.13	0.69	1	0.17	-0.41	-0.13	0.045	0.02
speechiness	0.25	0.14	0.17	1	-0.16	-0.056	0.077	-0.022
acousticness	-0.31	-0.59	-0.41	-0.16	1	0.017	0.039	-0.12
instrumentalness	0.0048	0.003	-0.13	-0.056	0.017	1	-0.012	0.049
liveness	-0.13	0.11	0.045	0.077	0.039	-0.012	1	0.023
valence	0.4	0.35	0.02	-0.022	-0.12	0.049	0.023	1

Results

- How has the sentiment of lyrics in popular music changed over time?

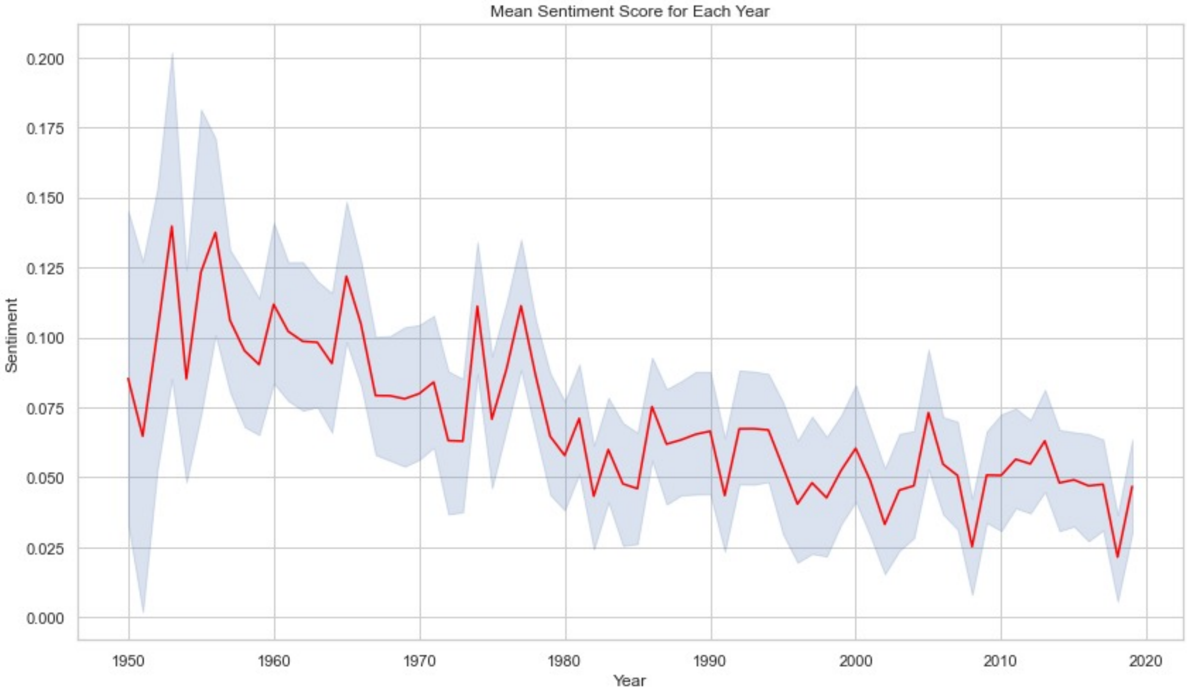
Successful Songs:



Results

- How has the sentiment of lyrics in popular music changed over time?

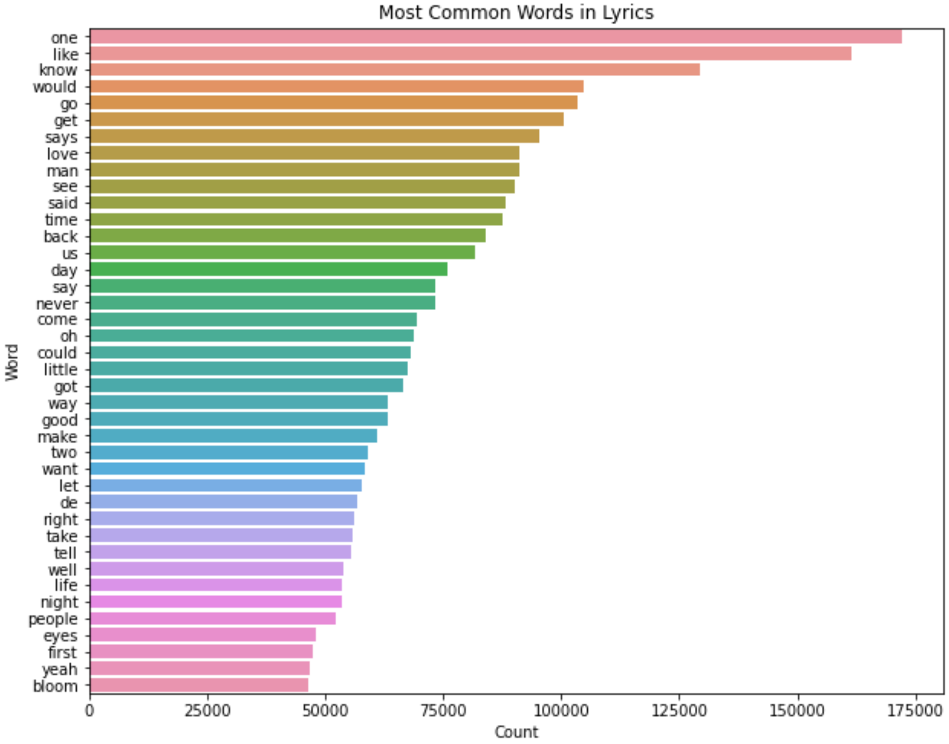
Unsuccessful Songs:



Results

- What are the top keywords in successful song lyrics over time?

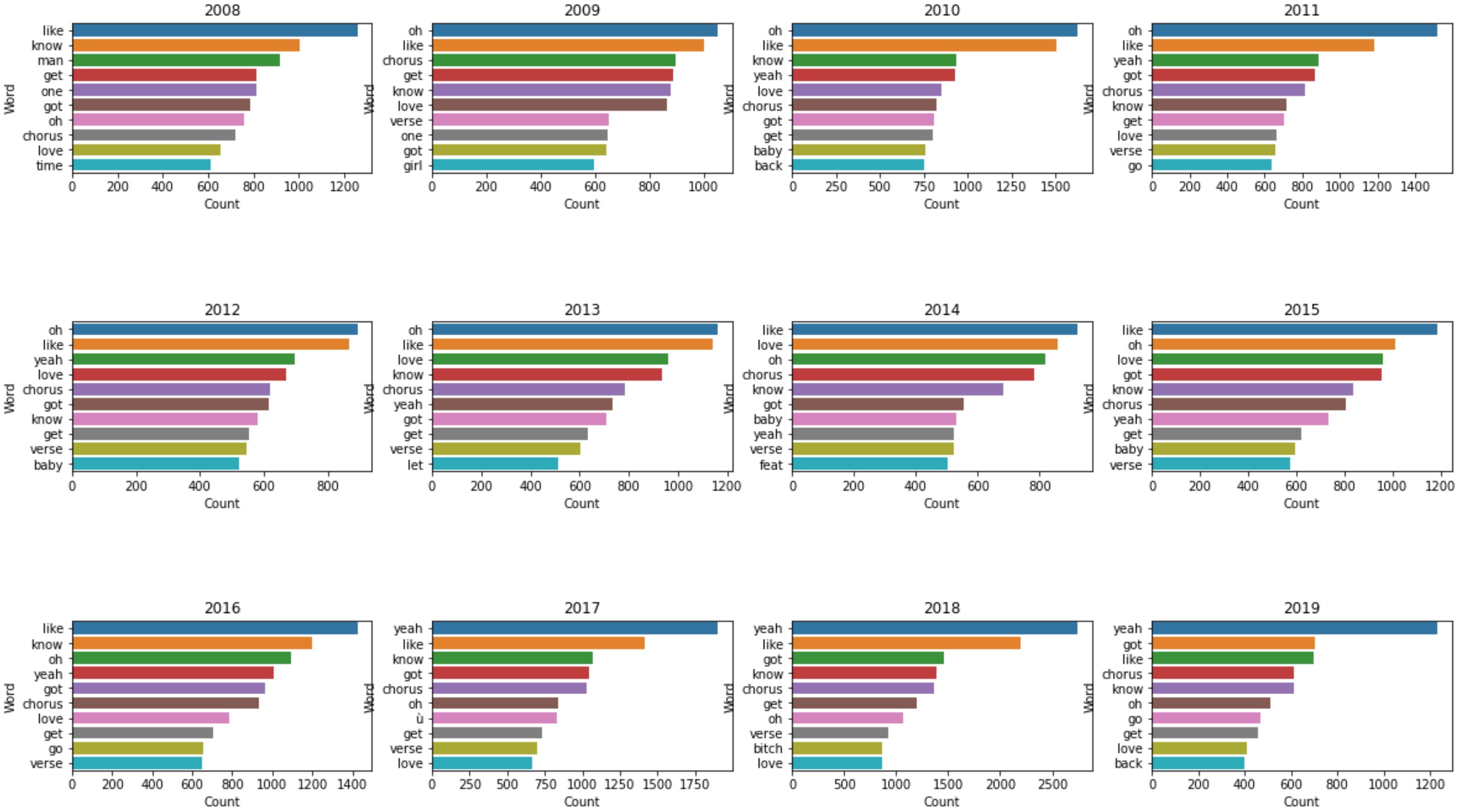
Successful Songs:



Results

- What are the top keywords in successful song lyrics over time?

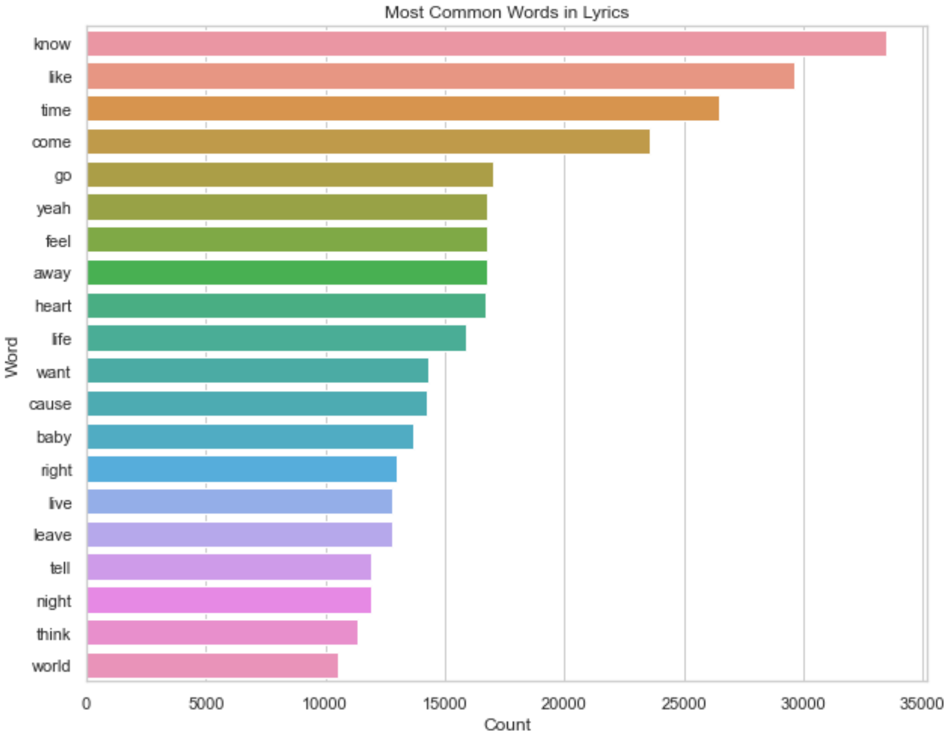
Successful Songs:



Results

- What are the top keywords in successful song lyrics over time?

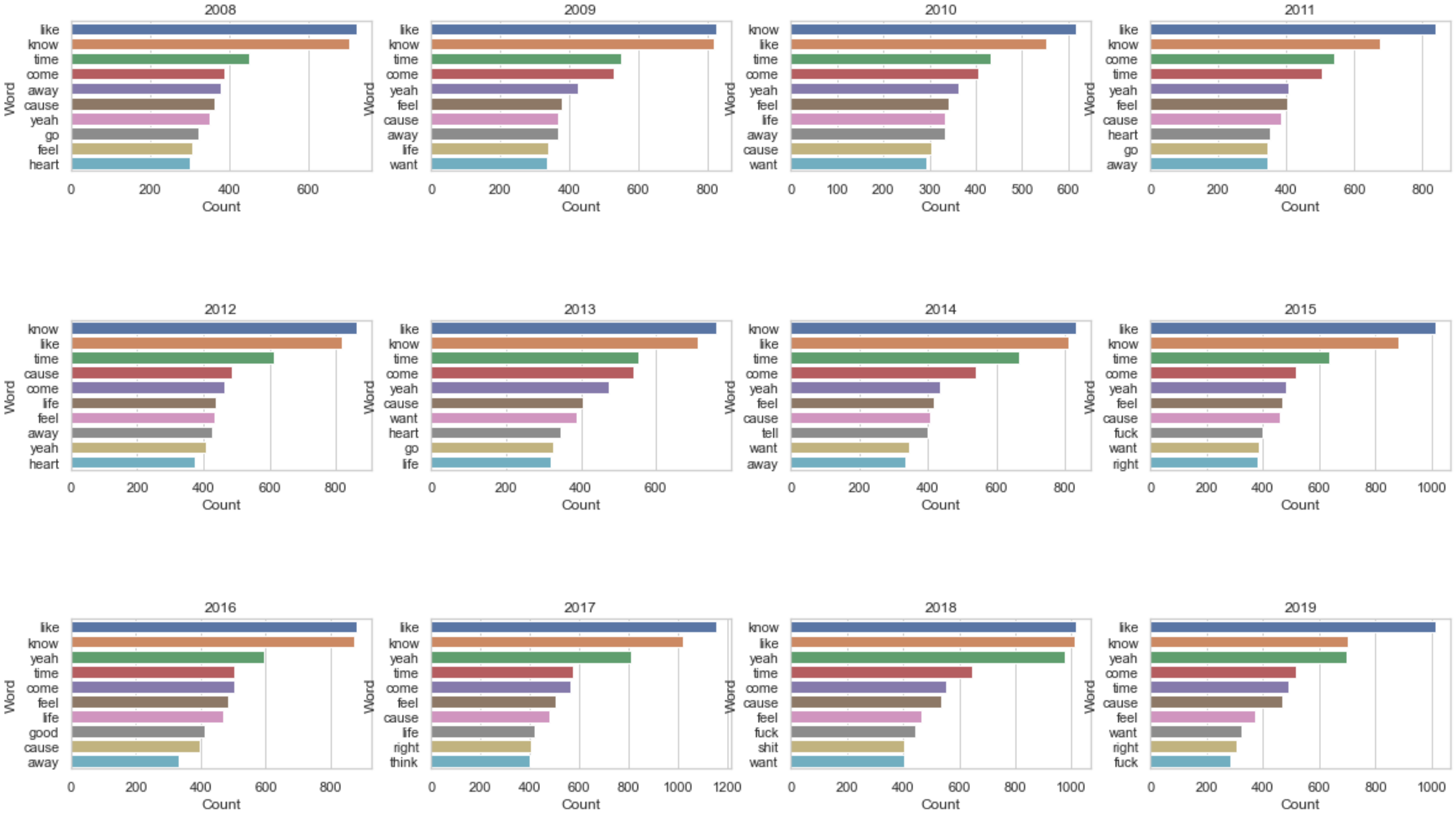
Unsuccessful Songs:



Results

- What are the top keywords in successful song lyrics over time?

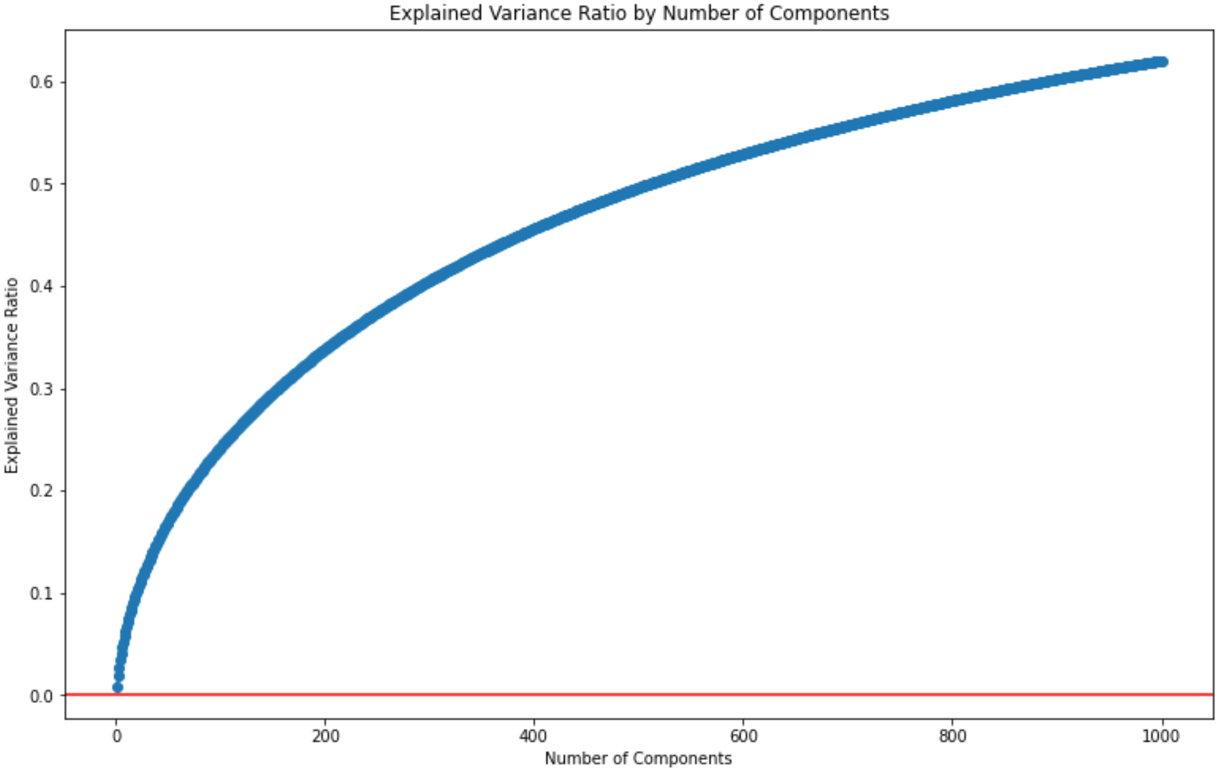
Unsuccessful Songs:



Results

- Song lyrics' variance can be described using how many components?

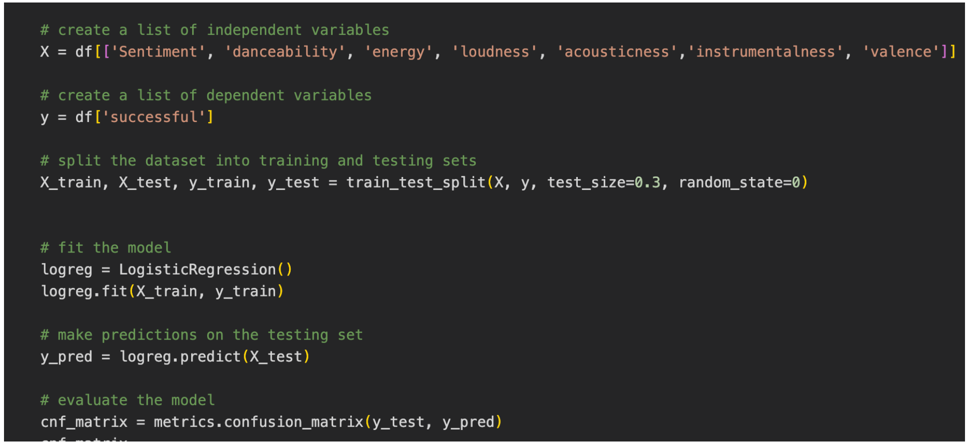
Successful Songs:



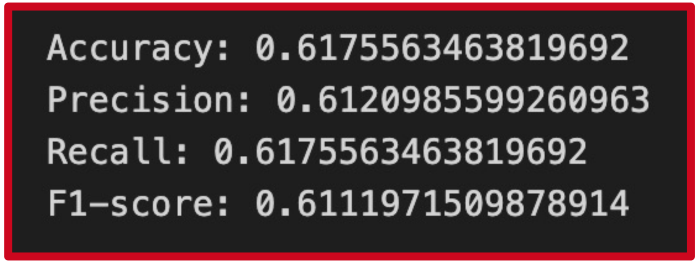
Reflection

Reflection

- I performed a logistic regression analysis on the set of all songs (successful and unsuccessful) to assess the effect of the different variables (e.g., sentiment) on the probability that a song becomes a hit song.

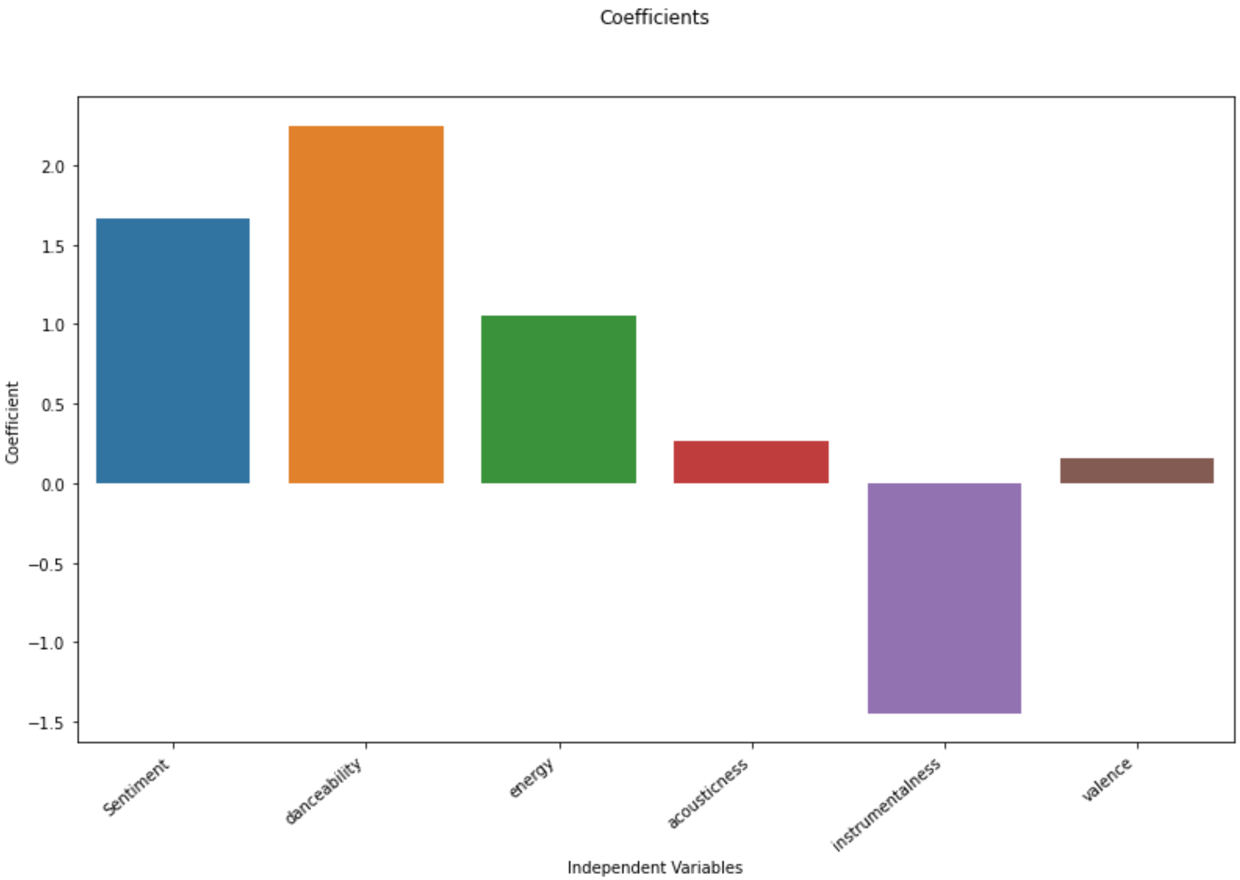


```
# create a list of independent variables
X = df[['Sentiment', 'danceability', 'energy', 'loudness', 'acousticness','instrumentalness', 'valence']]

# create a list of dependent variables
y = df['successful']

# split the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)


# fit the model
logreg = LogisticRegression()
logreg.fit(X_train, y_train)

# make predictions on the testing set
y_pred = logreg.predict(X_test)

# evaluate the model
cnf_matrix = metrics.confusion_matrix(y_test, y_pred)
cnf_matrix
```



```
Accuracy: 0.6175563463819692
Precision: 0.6120985599260963
Recall: 0.6175563463819692
F1-score: 0.6111971509878914
```

Reflection


Coefficients
Coefficient
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
Sentiment
danceability
energy
acousticness
instrumentalness
valence
Independent Variables

Limitations

Limitations

Weaknesses

- **Time period limitations:** Sample does not include 4 years recent songs.
- **Generalizability:** The findings of your research may be specific to the Billboard Hot 100 chart and the U.S. market. It might not necessarily apply to other charts or international markets, which may have different criteria and preferences for success.
- **Theoretical limitations:** research questions cannot be answered by empirical data alone without theoretical frameworks for guidance.

Future Work

Future Work

Future Steps:

- **Expand the dataset:** Consider incorporating a larger and more diverse dataset that includes songs beyond the Billboard Hot 100 chart.
- **Cross-cultural analysis:** Explore the variations in success factors and trends across different cultural contexts. This could involve comparing the Billboard Hot 100 chart with charts from other countries or analyzing songs from different cultural backgrounds.
- **Machine learning-based prediction models:** Develop predictive models using machine learning techniques to forecast the potential success of a song based on its features, sentiment, and keywords.

Thank you for your attention!

Sayedmahdi.raghib@rwth-aachen.de