beatles-lyrics-analysis

Background

Ahead of the 60th anniversary of the Beatles debut album Please Please Me being released, Company A has been commissioned by a media agency to do an analysis of the lyrics of the Beatles song catalogue. To add even more excitement 5 previously unreleased songs have been discovered and will be available for fans to listen to for the first time.

They are particularly interested in the themes of the Beatles songs, and if they can be grouped together into clusters of songs. They believe this will be important so they can quickly discover what other songs the new songs should be compared to - whether it be peace, love, or any other topic.

As the media agency does some kind of celebration every few years, they have already collected all the lyrics for me and supplied them as text files. A quick look indicates that there may be duplicates in the files where there are additional live recordings that have the same lyrics, and that some files might have weird or missing data - the media company has mentioned that if anything looks amiss then it should be removed.

The media company can be very fussy, so they expect to see a small powerpoint deck outlining the results, and justifications and explanations of any methodology that is used.

Data

- 279 files in a "lyrics" folder that are original lyrics by the Beatles to be analysed
- 5 files in a "new lyrics" folder, that are new songs by the Beatles soon to be released

Data Extraction

Folder with txt files -> A DataFrame with columns of songs names and lyrics

	Songs	Lyrics
0	Anna-go-to-him	Anna, you come and ask me girl\nTo set you fre...
1	yesterday	Yesterday, all my troubles seemed so far away\...
2	i-need-you	You don't realize how much I need you\nLove yo...
3	the-ballad-of-john-and-yoko	Standing in the dock at Southampton\ntrying to...
4	one-after-909	My baby says she's traveling\non the one after...

Data Cleaning

- Removed blank files and songs with duplicated lyrics.
- Number of songs before cleaning: 279
- Number of songs after cleaning: 217
- Number of blank files: 58
- Number of duplicated songs: 4
- Total number of songs removed: 62

Exploratory Data Analysis (EDA)

How many of the songs feature at least one pair of lines that rhyme?

- **Define rhyme:** the last words in the consecutive two lines of lyrics are ending with the same 2 letters.
- **Method:** For each song's lyrics, split into lines and extract the last words from every two consecutive lines and check whether the two letters in the last words match.
- **Data preprocessing:** Lowercase the lyrics.
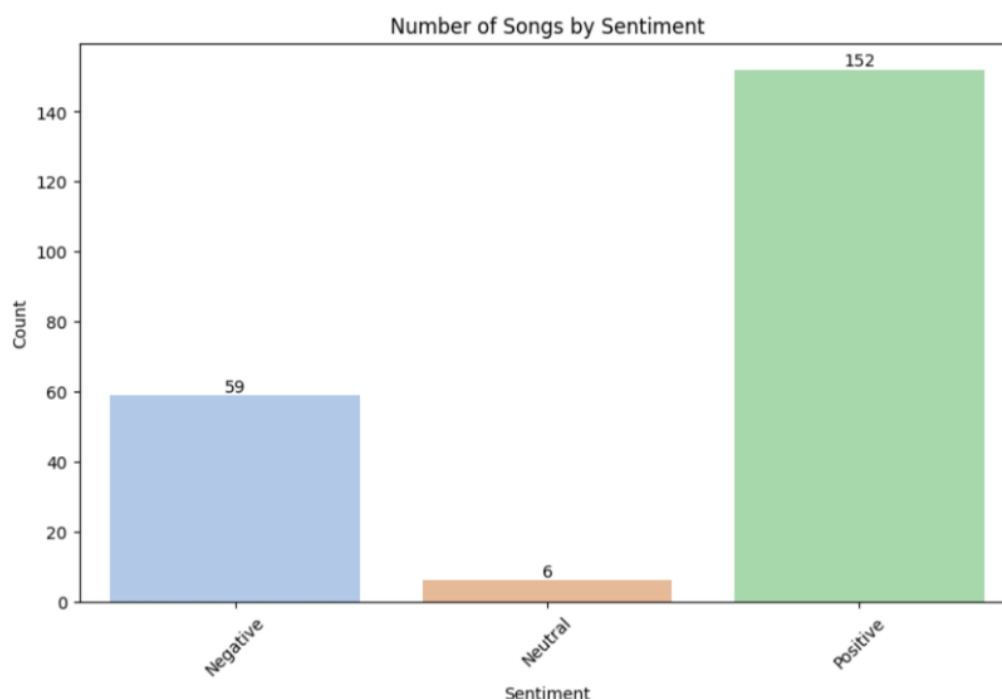- Number of songs with at least one pair of rhyming lines: 199

Cleaned_Songs	Cleaned_lyrics	Song Name in Lyrics
yesterday	yesterday all my troubles seemed so far away n...	True
i need you	you dont realize how much i need you love you ...	True
one after 909	my baby says shes traveling on the one after 9..	True
shes a woman	my love dont give me presents i know that shes...	True
three cool cats	three cool cats three cool cats are coming up ...	True

How many of the songs feature the song name (found in the file name) in the song lyrics?

- **Method:** For each song, search in its lyrics whether the song name is in it.
- **Data Preprocessing:** The dash in song names are replaced with space. The lyrics are lowercased, removed whitespace, and punctuations.
- Number of songs that feature the song name in the lyrics: 186

Sentiment Analysis

- **Method:** Used SentimentIntensityAnalyzer to score the lyrics and classify them as "Negative", "Neutral", or "Positive" (using +-0.05 thresholds).
- **Data Preprocessing:** Further cleaned lyrics by tokenizing and removing stopwords.

Negative Sentiment

Neutral Sentiment

Positive Sentiment

Topic Clustering

Method: Latent Dirichlet Allocation (LDA) topic modeling Data Preprocessing:

- Lemmatization
- Create a dictionary representation of the documents
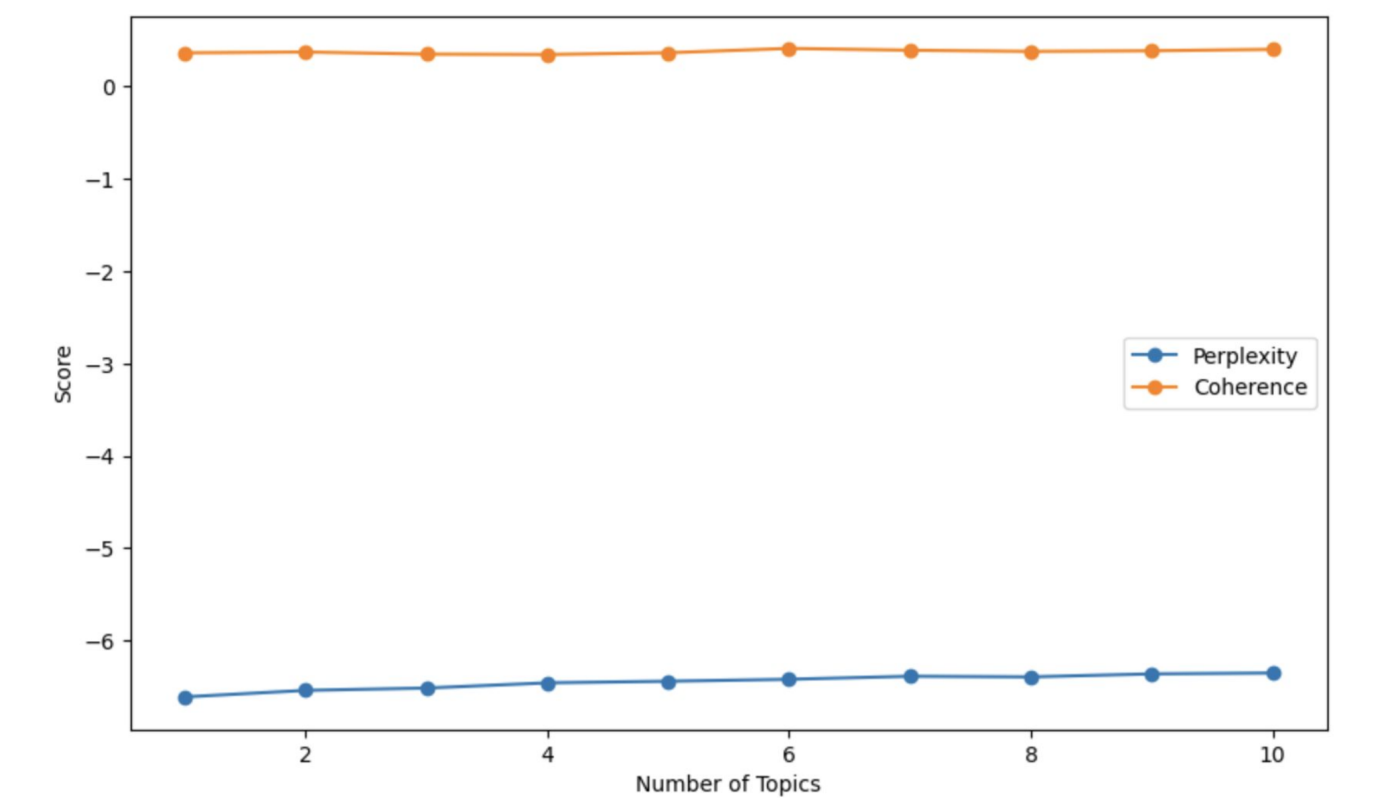- Convert the tokenized documents into a Document-Term Matrix (DTM)

Number of topics: 5 (according to the Perplexity and Coherence scores)

Choose number of topics

Lower perplexity values indicate a better fit to the data. Higher coherence scores indicate that the topics are more interpretable and coherent. In our case, 5 or 6 seems suitable.

- 5 topics: It has a relatively low perplexity score and a good coherence score.
- 6 topics: It has a slightly higher perplexity score but the highest coherence score, indicating more interpretable topics.

```
Nmber of topics: 1, Perplexity score: -6.607932699363433, Coherence score: 0.35940413508114166
Nmber of topics: 2, Perplexity score: -6.53728265365617, Coherence score: 0.36852994107013926
Nmber of topics: 3, Perplexity score: -6.512135686049541, Coherence score: 0.34551849415797603
Nmber of topics: 4, Perplexity score: -6.455839594386056, Coherence score: 0.34144181985679906
Nmber of topics: 5, Perplexity score: -6.438364191590927, Coherence score: 0.3603690789544701
Nmber of topics: 6, Perplexity score: -6.417471866162751, Coherence score: 0.40753132169080336
Nmber of topics: 7, Perplexity score: -6.384586508711286, Coherence score: 0.38761994315402554
Nmber of topics: 8, Perplexity score: -6.391590230861351, Coherence score: 0.3750698240007032
Nmber of topics: 9, Perplexity score: -6.358783049099269, Coherence score: 0.3817416989714838
Nmber of topics: 10, Perplexity score: -6.349536488604738, Coherence score: 0.39694951816645435
```


Interpretation of Clusters

```
Topic 0: 0.050*"love" + 0.024*"im" + 0.021*"know" + 0.016*"want" + 0.015*"got" + 0.013*"ill" + 0.013*"oh" + 0.013*"girl" + 0.011*"dont" + 0.011*"yeah"
Topic 1: 0.027*"na" + 0.014*"let" + 0.013*"gon" + 0.013*"see" + 0.013*"well" + 0.013*"im" + 0.012*"oh" + 0.011*"dont" + 0.010*"together" + 0.010*"said"
Topic 2: 0.030*"dont" + 0.021*"back" + 0.015*"na" + 0.014*"know" + 0.013*"get" + 0.012*"say" + 0.012*"im" + 0.012*"wont" + 0.012*"man" + 0.011*"come"
Topic 3: 0.037*"know" + 0.029*"yeah" + 0.019*"baby" + 0.015*"oh" + 0.014*"want" + 0.013*"da" + 0.013*"right" + 0.013*"well" + 0.012*"ah" + 0.010*"bill"
Topic 4: 0.016*"going" + 0.014*"come" + 0.014*"cry" + 0.013*"lonely" + 0.013*"way" + 0.013*"mine" + 0.012*"take" + 0.011*"shes" + 0.011*"youre" + 0.010*"roll"
```

Topic 1: Love and Affection

This topic seems to revolve around themes of love and affection. Words like "love," "im," "know," "want," "got," and "girl" suggest a strong emphasis on love and romantic relationships. It conveys feelings of desire and affection.

Topic 2: Actions and Expressions

This topic appears to be about actions and expressions. Words like "na," "let," "gon," "see," and "together" suggest a focus on taking action and expressing oneself. It might involve planning or making decisions.
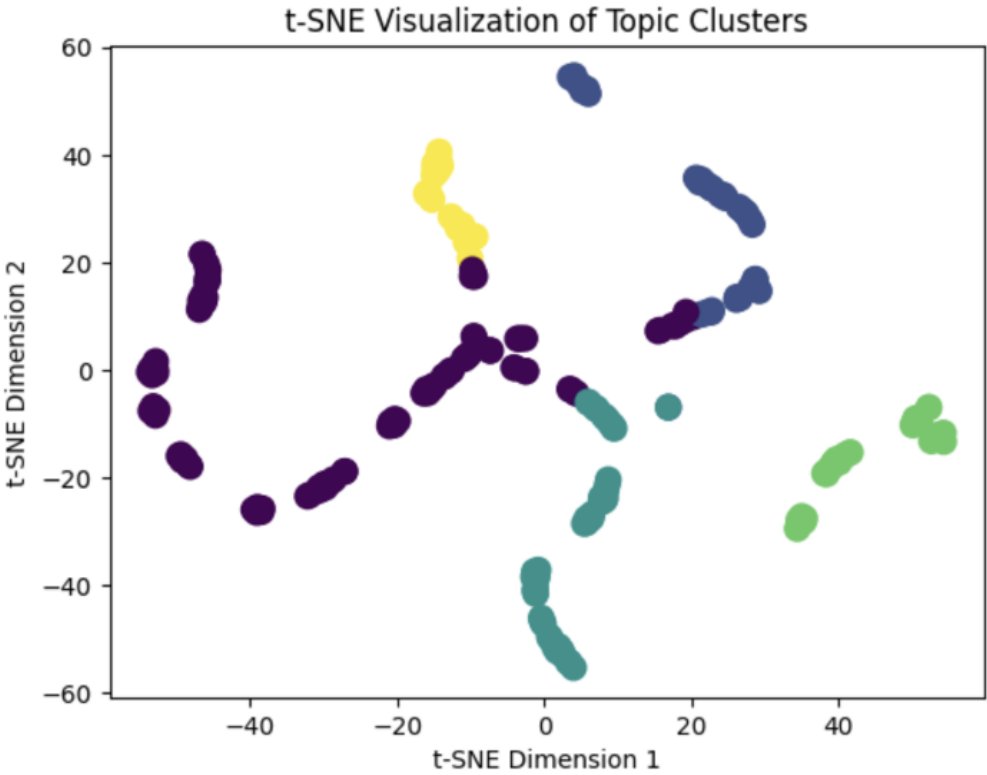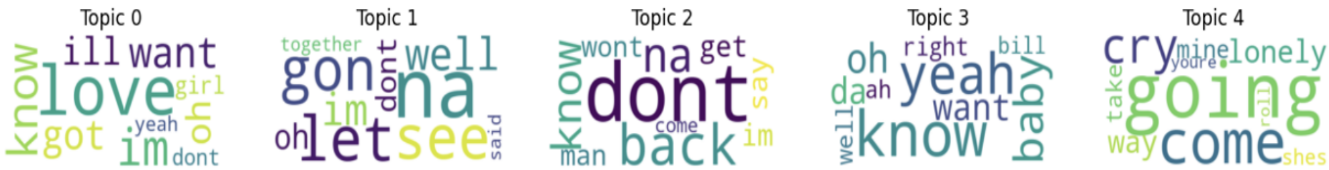
Topic 3: Assertiveness and Statements

This topic seems to involve assertiveness and statements. Words like "dont," "back," "na," "get," "say," and "wont" indicate a tone of making statements or expressing one's perspective with confidence.

Topic 4: Positive Emotions and Affirmations

This topic is related to positive emotions and affirmations. Words like "know," "yeah," "baby," "oh," and "want" indicate a positive and enthusiastic tone. It might involve expressions of joy, affirmation, or optimism.

Topic 5: Emotions and Expression

This topic could be about emotions and expressions. Words like "going," "come," "cry," "lonely," and "mine" suggest a focus on emotions and expressions, possibly related to personal experiences and feelings.

Result

```
                  Songs  Predicted Topic
     the-sunshine-arrived                0
     peace-love-and-peace                0
a-very-beatles-christmas                3
love-is-all-that-s-needed                0
   a-less-yellow-submarine                1
```

- Topic 0: Love and Affection
- Topic 1: Actions and Expressions
- Topic 2: Assertiveness and Statements
- Topic 3: Positive Emotions and Affirmations
- Topic 4: Emotions and Expression

Result

Topic 0: Love and Affection

The sunshine arrived:

In the morning light I feel so free
With the sun shining bright, just you and me
Walking hand in hand, down a road of dreams
With every step we take, it seems like we're living in a scene
Oh, we're walking on sunshine, and the sky is so blue
With the breeze blowing softly, it feels like we're flying too
And the world is so bright, when we're together, it's true
Oh, we're walking on sunshine, living life like it's new
In the fields of gold, we laugh and we play
With the flowers and the trees, it feels like a perfect day
We dance and we sing, to the rhythm of our hearts
With every beat and every note, it feels like we're never apart
Oh, we're walking on sunshine, and the sky is so blue
With the breeze blowing softly, it feels like we're flying too
And the world is so bright, when we're together, it's true
Oh, we're walking on sunshine, living life like it's new
With every smile and every kiss, it feels like a dream come true
And every moment we spend, together it's all brand new
Oh, we're walking on sunshine, and the sky is so blue
With the breeze blowing softly, it feels like we're flying too
And the world is so bright, when we're together, it's true
Oh, we're walking on sunshine, living life like it's new
And when the night falls and the stars shine bright
We'll be walking on sunshine, into the endless night.

Result

Topic 0: Love and Affection

Peace love and peace:

In a world that's filled with chaos and strife
We search for a beacon, to lead us to life
With love in our hearts, we'll conquer all fear
With peace in our minds, we'll spread it far and near
Peace and love, it's what we all need
With these words, we'll plant the seeds
Of a world that's filled with harmony and glee
Where everyone can live in peace and love, you'll see
From the mountains high, to the valleys below
We'll sing out this message, for all to know
That peace and love, is what we need the most
And together we'll build a world that's close
Peace and love, it's what we all need
With these words, we'll plant the seeds
Of a world that's filled with harmony and glee
Where everyone can live in peace and love, you'll see
Let us spread this message, to every land and sea
With peace and love, we'll set our spirits free
Peace and love, it's what we all need
With these words, we'll plant the seeds
Of a world that's filled with harmony and glee
Where everyone can live in peace and love, you'll see
With peace and love, we'll make a brighter day
In a world that's filled with peace and love, we'll find our way.

Result

Topic 3: Positive Emotions and Affirmations

A very beatles christmas:

It's that time of year, when the world comes alive
With the sounds of joy, and the twinkling of lights
The air is so crisp, and the snow falls down
And the spirit of Christmas, can be felt all around
It's the most wonderful time of the year
With the laughter and cheer, ringing in our ears
It's a season of love, and a season of peace
And a time to be grateful, for all that we share and believe
In the homes and the streets, the carolers sing
With the songs of the season, their voices take wing
And the children are giddy, with excitement and glee
As they dream of the gifts, under the tree
It's the most wonderful time of the year
With the laughter and cheer, ringing in our ears
It's a season of love, and a season of peace
And a time to be grateful, for all that we share and believe
Let us spread love, and kindness all around
And make the world brighter, with love that we've found
It's the most wonderful time of the year
With the laughter and cheer, ringing in our ears
It's a season of love, and a season of peace
And a time to be grateful, for all that we share and believe
So let's embrace the season, with open hearts and smiles
And make this Christmas, the best one in miles!

Result

Topic 0: Love and Affection

Love is all that's needed:

In a world that's so complex, and filled with pain
We search for something, that can ease our strain
And as we journey through life, with every step we take
We find that the answer, is love that we make
All we need is love, to light up the way
With a heart that's so pure, we'll find peace each day
With love in our hearts, we'll conquer all fear
And we'll know that in this world, love is all we need here
With every smile, and every gentle touch
Love has a way, of healing us so much
And as we give love, it returns tenfold
With love in our lives, we'll never grow old
All we need is love, to light up the way
With a heart that's so pure, we'll find peace each day
With love in our hearts, we'll conquer all fear
And we'll know that in this world, love is all we need here
Let us spread love, and kindness all around
And make the world brighter, with love that we've found
All we need is love, to light up the way
With a heart that's so pure, we'll find peace each day
With love in our hearts, we'll conquer all fear
And we'll know that in this world, love is all we need here
So let's embrace love, with open hearts and smiles
And know that in this world, love is all we need to find!

Result

Topic 1: Actions and Expressions

A less yellow submarine:

Sailing under the sea, in a metal machine
Exploring the unknown, in the deep blue marine
Through the currents and waves, we journey on
In this submarine, we'll never be alone
Diving in the ocean, exploring all we can see
With the submarines, we'll conquer the deep blue sea
With every dive, we discover new mysteries
In the underwater world, we'll make our own histories
Gliding through the depths, with lights shining bright
With every mile we travel, we're venturing into the night
Watching schools of fish, and creatures of the sea
In this metal vessel, we're living so carefree
Diving in the ocean, exploring all we can see
With the submarines, we'll conquer the deep blue sea
With every dive, we discover new mysteries
In the underwater world, we'll make our own histories
Through the endless ocean, we'll journey on and on
With submarines, we'll never be gone
Diving in the ocean, exploring all we can see
With the submarines, we'll conquer the deep blue sea
With every dive, we discover new mysteries
In the underwater world, we'll make our own histories
And when the mission's done, and we return to shore
We'll remember all the wonders, that the ocean has in store.

Limitations

- Fixed number of topics
- Bag of words
- Lack of Sufficient Data
- Limited Vocabulary
- Unstable Results
- Difficulty in Interpreting Topics

Future Work

- Hyperparameter tuning
- Exploring alternative topic modeling techniques that may be more suitable for small datasets
- Combine LDA with other clustering methods