TaylorSwift Lyrics Analysis using Sentimental Analysis

The objective of this project is to gain a deeper understanding of the recurring themes in Taylor Swift's songs and how these themes have evolved throughout her career.This code allows for a data-driven exploration of the lyrical content,providing valuable insights into the artist's artistic expression and creative journey.

In [1]:
```
# Import the pandas library as pd
import pandas as pd
```

In [2]:
```
# Read the CSV file containing Taylor Swift's song lyrics into a DataFram
df = pd.read_csv('/Users/user/Downloads/taylor_swift_lyrics.csv', encodin
```

In [3]:
```
# Display the first few rows of the DataFrame 'df' using the head() funct
df.head()
```

Out[3]:

	artist	album	track_title	track_n	lyric	line	year
0	Taylor Swift	Taylor Swift	Tim McGraw	1	He said the way my blue eyes shined	1	2006
1	Taylor Swift	Taylor Swift	Tim McGraw	1	Put those Georgia stars to shame that night	2	2006
2	Taylor Swift	Taylor Swift	Tim McGraw	1	I said, "That's a lie"	3	2006
3	Taylor Swift	Taylor Swift	Tim McGraw	1	Just a boy in a Chevy truck	4	2006
4	Taylor Swift	Taylor Swift	Tim McGraw	1	That had a tendency of gettin' stuck	5	2006

In [4]:
```
# Display the last few rows of the DataFrame 'df' using the tail() functi
df.tail()
```

Out[4]:

	artist	album	track_title	track_n	lyric	line	year
4857	Taylor Swift	reputation	New Year's Day	15	(Hold on to the memories, they will hold on to...	43	2017
4858	Taylor Swift	reputation	New Year's Day	15	Please don't ever become a stranger	44	2017
4859	Taylor Swift	reputation	New Year's Day	15	(Hold on to the memories, they will hold on to...	45	2017
4860	Taylor Swift	reputation	New Year's Day	15	Whose laugh I could recognize anywhere	46	2017
4861	Taylor Swift	reputation	New Year's Day	15	(I will hold on to you)	47	2017

In [5]:
```
num_rows = df.shape[0]
num_columns = df.shape[1]
print("Number of rows:", num_rows)
print("Number of columns:", num_columns)
```

```
Number of rows: 4862
Number of columns: 7
```

In [6]:
```
column_names = df.columns
print(column_names)
```

```
Index(['artist', 'album', 'track_title', 'track_n', 'lyric', 'line', 'yea
r'], dtype='object')
```

In [7]:
```
data_types = df.dtypes
print(data_types)
```

```
artist         object
album          object
track_title    object
track_n         int64
lyric          object
line            int64
year            int64
dtype: object
```

In [8]:
```
# Check for missing values in the DataFrame
missing_values = df.isnull().sum()

# Print the number of missing values for each column
print(missing_values)
```

```
artist         0
album          0
track_title    0
track_n        0
lyric          0
line           0
year           0
dtype: int64
```

In [9]:
```
summary_statistics = df.describe()
print(summary_statistics)
```

```
           track_n         line         year
count  4862.000000  4862.000000  4862.000000
mean      8.216989    28.426573  2011.882764
std       4.696379    18.343649     3.571447
min       1.000000     1.000000  2006.000000
25%       4.000000    13.000000  2010.000000
50%       8.000000    26.000000  2012.000000
75%      12.000000    41.000000  2014.000000
max      19.000000   101.000000  2017.000000
```

In [10]:

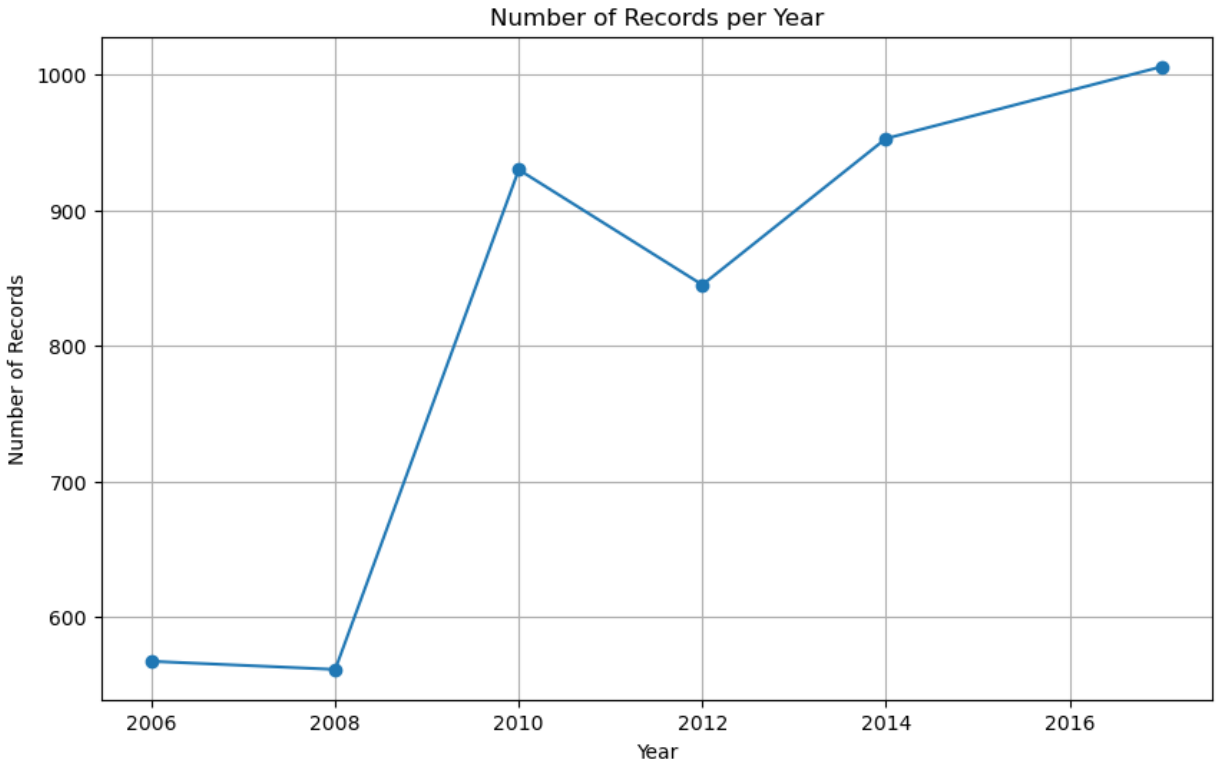
```python
import pandas as pd
# Distribution of track_n values
track_n_distribution = df['track_n'].value_counts()

# Distribution of line values
line_distribution = df['line'].value_counts()

print("Distribution of track_n values:")
print(track_n_distribution)

print("\nDistribution of line values:")
print(line_distribution)
```

```
Distribution of track_n values:
2     368
7     338
4     337
6     333
9     331
8     330
1     315
5     311
10    307
3     284
14    283
11    279
13    265
12    259
15    177
16    144
17    116
19     47
18     38
Name: track_n, dtype: int64

Distribution of line values:
1      94
2      94
22     94
21     94
20     94
      ..
87      1
86      1
85      1
84      1
101     1
Name: line, Length: 101, dtype: int64
```

In [11]:

```python
import pandas as pd
import matplotlib.pyplot as plt

# Year Statistics
average_year = df['year'].mean()
earliest_year = df['year'].min()
latest_year = df['year'].max()

print("Average Year:", average_year)
print("Earliest Year:", earliest_year)
print("Latest Year:", latest_year)

# Number of records per year
records_per_year = df['year'].value_counts().sort_index()

print("\nNumber of Records per Year:")
print(records_per_year)

# Trends and Patterns in the year column (Plotting)
plt.figure(figsize=(10, 6))
plt.plot(records_per_year.index, records_per_year.values, marker='o')
plt.xlabel('Year')
plt.ylabel('Number of Records')
plt.title('Number of Records per Year')
plt.grid(True)
plt.show()
```

```
Average Year: 2011.882764294529
Earliest Year: 2006
Latest Year: 2017

Number of Records per Year:
2006     567
2008     561
2010     930
2012     845
2014     953
2017    1006
Name: year, dtype: int64
```



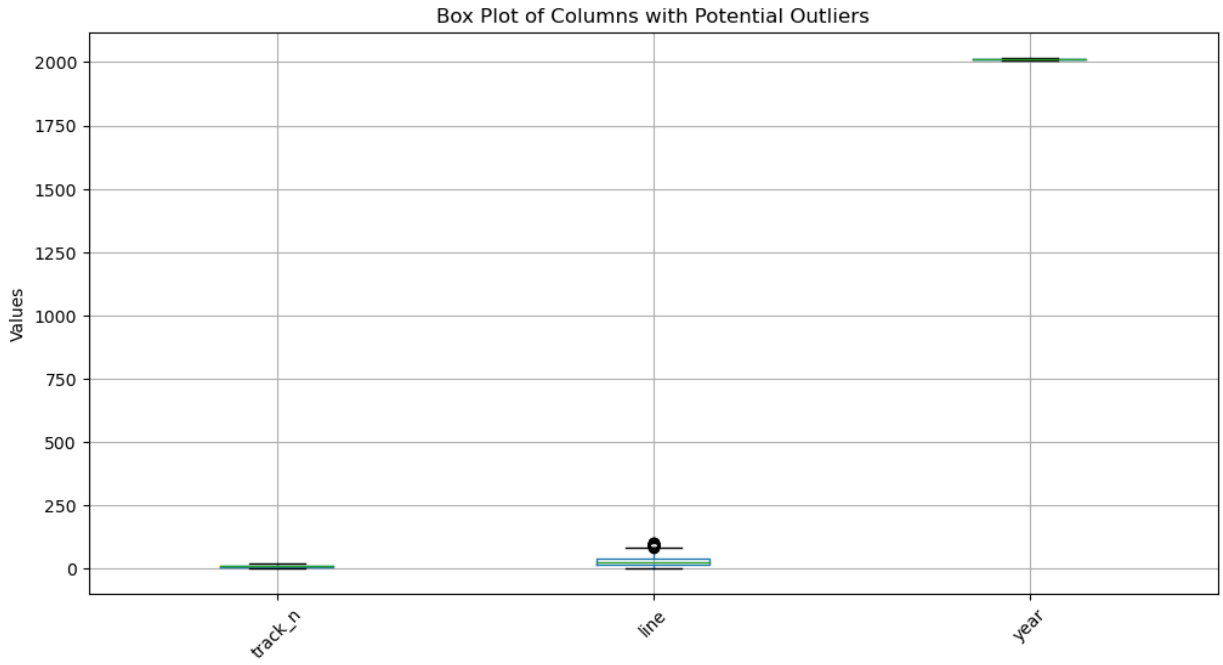
In [12]:
```
# Variation of track_n
track_n_std = df['track_n'].std()

# Variation of line
line_std = df['line'].std()

print("Standard Deviation of track_n:", track_n_std)
print("Standard Deviation of line:", line_std)
```

```
Standard Deviation of track_n: 4.696378535310443
Standard Deviation of line: 18.343649158618703
```

In [13]:
```
# Earliest and Latest Years
earliest_year = df['year'].min()
latest_year = df['year'].max()

# Minimum and Maximum values for track_n and line columns
min_track_n = df['track_n'].min()
max_track_n = df['track_n'].max()

min_line = df['line'].min()
max_line = df['line'].max()

print("Earliest Year:", earliest_year)
print("Latest Year:", latest_year)

print("Minimum Track_n:", min_track_n)
print("Maximum Track_n:", max_track_n)

print("Minimum Line:", min_line)
print("Maximum Line:", max_line)
```

```
Earliest Year: 2006
Latest Year: 2017
Minimum Track_n: 1
Maximum Track_n: 19
Minimum Line: 1
Maximum Line: 101
```

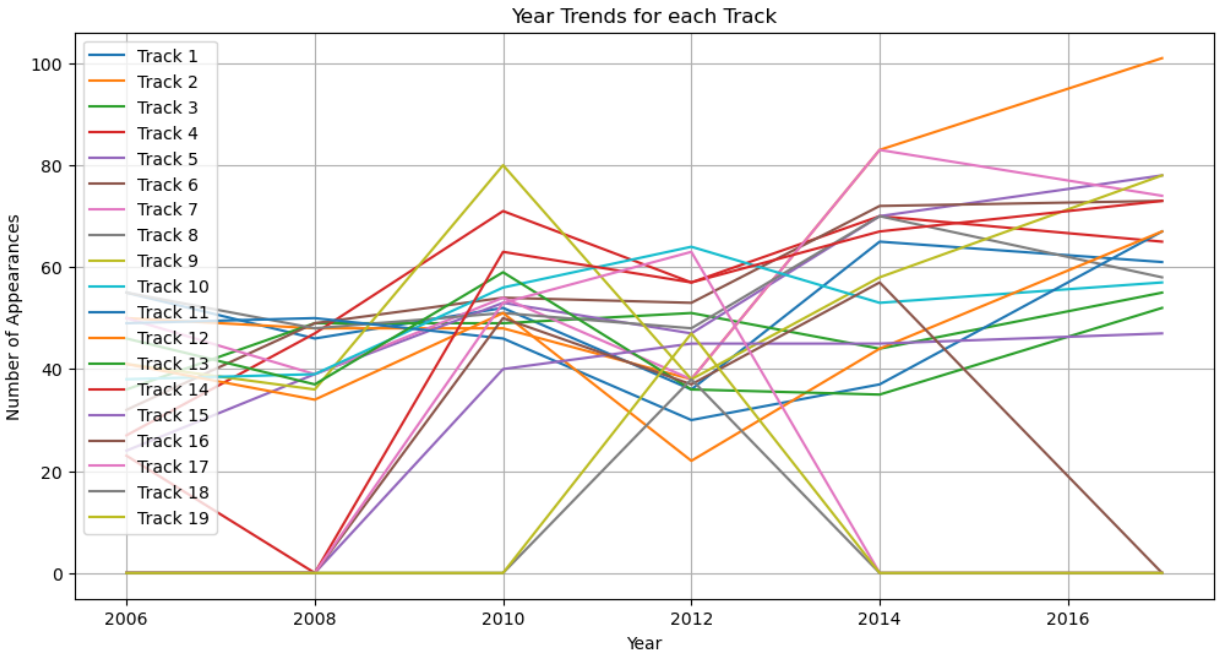
In [14]:
```
# Columns to check for outliers
columns_to_check = ['track_n', 'line', 'year']

# Outlier detection using box plots
plt.figure(figsize=(12, 6))
df[columns_to_check].boxplot()
plt.title('Box Plot of Columns with Potential Outliers')
plt.ylabel('Values')
plt.xticks(rotation=45)
plt.show()
```

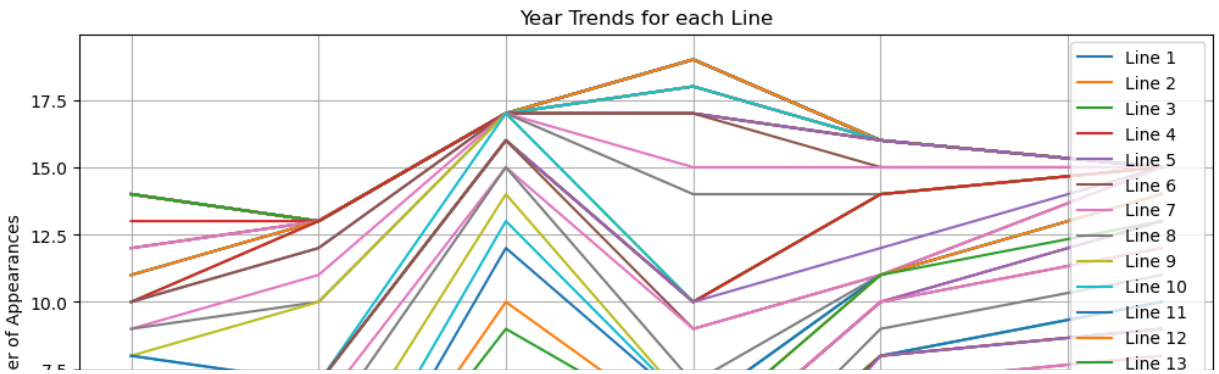


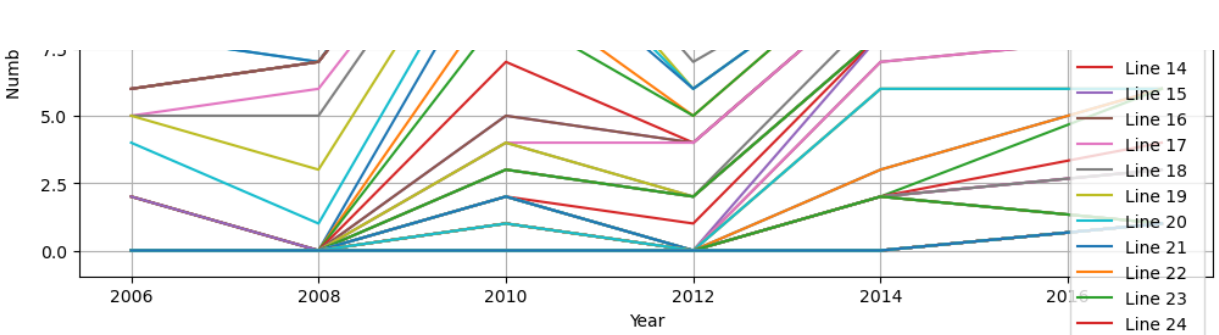
In [15]:
```
# Calculate correlation matrix
correlation_matrix = df[['track_n', 'line', 'year']].corr()

print("Correlation Matrix:")
print(correlation_matrix)
```

```
Correlation Matrix:
          track_n      line      year
track_n  1.000000 -0.089467  0.019301
line    -0.089467  1.000000  0.238740
year     0.019301  0.238740  1.000000
```

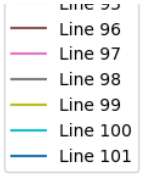
In [16]:
```
# Group the data by year and track_n to calculate the count of each track
year_track_count = df.groupby(['year', 'track_n']).size().unstack(fill_va

# Plotting Year Trends for each track_n
plt.figure(figsize=(12, 6))
for track in year_track_count.columns:
    plt.plot(year_track_count.index, year_track_count[track], label=f'Tra

plt.xlabel('Year')
plt.ylabel('Number of Appearances')
plt.title('Year Trends for each Track')
plt.legend()
plt.grid(True)
plt.show()

# Group the data by year and line to calculate the count of each line for
year_line_count = df.groupby(['year', 'line']).size().unstack(fill_value=

# Plotting Year Trends for each line
plt.figure(figsize=(12, 6))
for line in year_line_count.columns:
    plt.plot(year_line_count.index, year_line_count[line], label=f'Line {

plt.xlabel('Year')
plt.ylabel('Number of Appearances')
plt.title('Year Trends for each Line')
plt.legend()
plt.grid(True)
plt.show()
```






Numb
7.5
5.0
2.5
0.0
2006
2008
2010
2012
2014
2016
Year

Line 14
Line 15
Line 16
Line 17
Line 18
Line 19
Line 20
Line 21
Line 22
Line 23
Line 24
Line 25
Line 26
Line 27
Line 28
Line 29
Line 30
Line 31
Line 32
Line 33
Line 34
Line 35
Line 36
Line 37
Line 38
Line 39
Line 40
Line 41
Line 42
Line 43
Line 44
Line 45
Line 46
Line 47
Line 48
Line 49
Line 50
Line 51
Line 52
Line 53
Line 54
Line 55
Line 56
Line 57
Line 58
Line 59
Line 60
Line 61
Line 62
Line 63
Line 64
Line 65
Line 66
Line 67
Line 68
Line 69
Line 70
Line 71
Line 72
Line 73
Line 74
Line 75
Line 76
Line 77
Line 78
Line 79
Line 80
Line 81
Line 82
Line 83
Line 84
Line 85
Line 86
Line 87
Line 88
Line 89
Line 90
Line 91
Line 92
Line 93
Line 94



In [17]:
```python
# Calculate percentiles for track_n
track_n_percentiles = df['track_n'].quantile([0.25, 0.5, 0.75])

# Calculate percentiles for line
line_percentiles = df['line'].quantile([0.25, 0.5, 0.75])

# Calculate percentiles for year
year_percentiles = df['year'].quantile([0.25, 0.5, 0.75])

print("Track_n Percentiles:")
print(track_n_percentiles)

print("\nLine Percentiles:")
print(line_percentiles)

print("\nYear Percentiles:")
print(year_percentiles)
```

```
Track_n Percentiles:
0.25     4.0
0.50     8.0
0.75    12.0
Name: track_n, dtype: float64

Line Percentiles:
0.25    13.0
0.50    26.0
0.75    41.0
Name: line, dtype: float64

Year Percentiles:
0.25    2010.0
0.50    2012.0
0.75    2014.0
Name: year, dtype: float64
```

In [18]:
```python
# Get the count of songs for each year
songs_per_year = df['year'].value_counts()

print("Distribution of Songs Across Different Years:")
print(songs_per_year)
```

```
Distribution of Songs Across Different Years:
2017    1006
2014     953
2010     930
2012     845
2006     567
2008     561
Name: year, dtype: int64
```

In [19]:
```
import pandas as pd
from collections import Counter
import re


# Extract Taylor Swift's song lyrics from the 'lyric' column
taylor_lyrics = df[df['artist'] == 'Taylor Swift']['lyric']

# Preprocess the lyrics and count word frequencies
word_freq = Counter()

for lyrics in taylor_lyrics:
    # Convert lyrics to lowercase
    lyrics_lower = lyrics.lower()
    # Remove punctuation and split into words
    words = re.findall(r'\b\w+\b', lyrics_lower)
    # Update word frequencies
    word_freq.update(words)

# Get the top N most common words (e.g., top 10 words)
N = 10
most_common_words = word_freq.most_common(N)

print("Top", N, "most common words in Taylor Swift's song lyrics:")
for word, frequency in most_common_words:
    print(word, ":", frequency)
```

```
Top 10 most common words in Taylor Swift's song lyrics:
you : 2106
i : 1950
the : 1147
and : 1097
it : 828
to : 648
me : 637
a : 582
t : 569
in : 504
```

In [20]:
```
import pandas as pd
# Count the number of unique songs
unique_songs = df['track_title'].nunique()

# Count the number of unique artists
unique_artists = df['artist'].nunique()

print("Number of Unique Songs:", unique_songs)
print("Number of Unique Artists:", unique_artists)
```

```
Number of Unique Songs: 94
Number of Unique Artists: 1
```

In [21]:
```
# Sort the DataFrame based on the length of the 'lyric' column (ascending
sorted_df = df.sort_values(by='lyric', key=lambda x: x.str.len())

# Select the top 5 rows with the longest lyrics
top_longest_songs = sorted_df.tail(5)

# Select the top 5 rows with the shortest lyrics
top_shortest_songs = sorted_df.head(5)

print("Top 5 Songs with the Longest Lyrics:")
print(top_longest_songs[['track_title', 'lyric']])

print("\nTop 5 Songs with the Shortest Lyrics:")
print(top_shortest_songs[['track_title', 'lyric']])
```

```
Top 5 Songs with the Longest Lyrics:
          track_title                                              lyric
1501   The Story of Us  Yeah, and I don't know what to say since a twi...
1213        Sparks Fly  Just keep on keeping your eyes on me, it's jus...
4535  King of My Heart  And all at once, you are the one, I have been ...
1529     Never Grow Up  And don't lose the way that you dance around i...
3281      Shake It Off  And dirty, dirty cheats of the world you could...

Top 5 Songs with the Shortest Lyrics:
                            track_title lyric
2317                                 22    22
2331                                 22    22
2300                                 22    22
3740                         Wonderland   Oh!
3994  End Game (Ft. Ed Sheeran & Future)   ooh
```

In [22]:
```
# Define the word or phrase you want to search for
specific_word_or_phrase = 'love'  # Change this to your desired word or p

# Search for the word or phrase in the 'lyric' column
matching_songs = df[df['lyric'].str.contains(specific_word_or_phrase, cas

print("Songs with Lyrics Containing the Word or Phrase:", specific_word_o
print(matching_songs[['track_title', 'lyric']])
```

Songs with Lyrics Containing the Word or Phrase: love

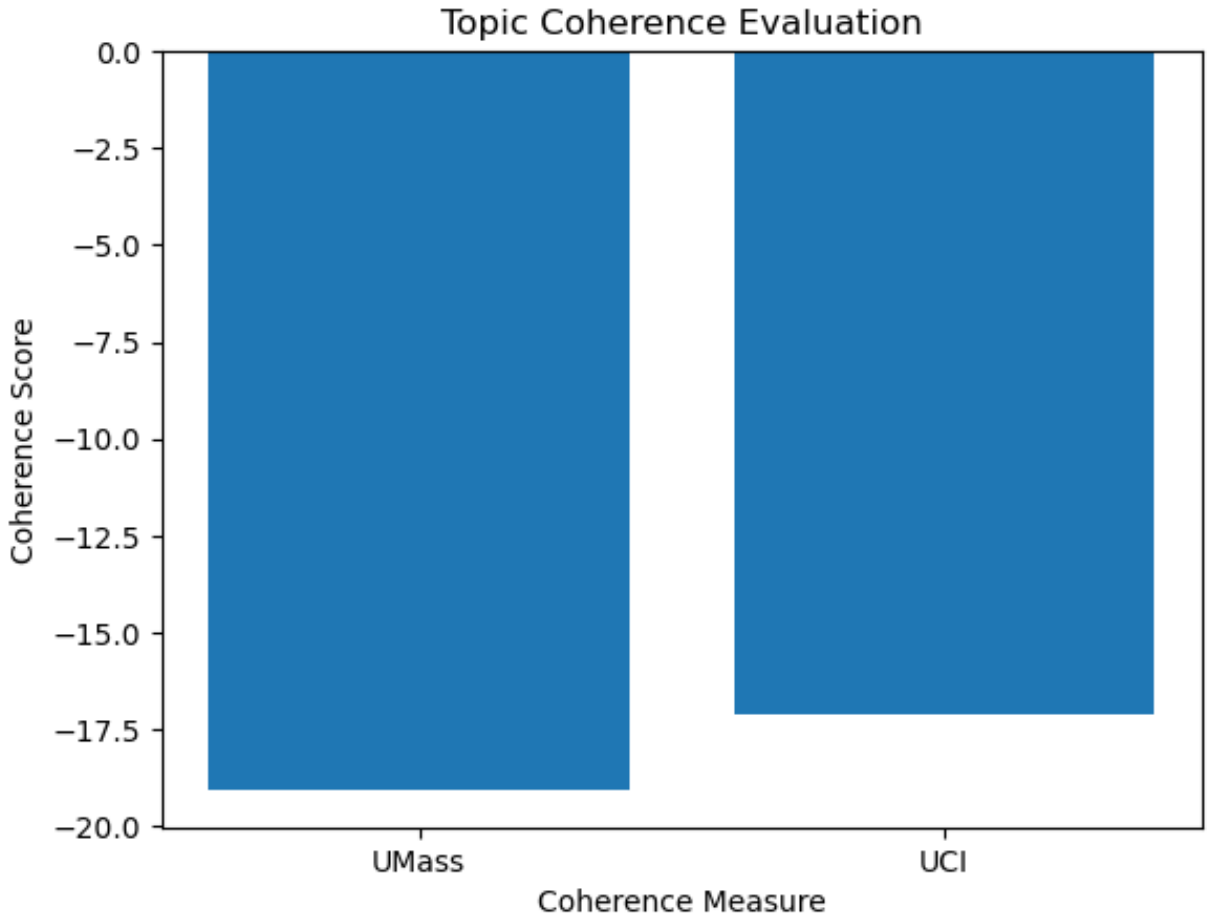
	track_title	lyric
56	Picture To Burn	I realize you love yourself more than you coul...
115	Teardrops On My Guitar	He says he's so in love, he's finally got it r...
125	Teardrops On My Guitar	She'd better hold him tight, give him all her ...
240	Tied Together With A Smile	I guess it's true that love was all you wanted
283	Stay Beautiful	Every little piece love
...	...	...
4581	Dancing With Our Hands Tied	I, I loved you in spite of
4756	Call It What You Want	Loves me like I'm brand new
4769	Call It What You Want	I'm laughing with my lover, makin' forts under...
4780	Call It What You Want	Loves me like I'm brand new
4802	Call It What You Want	Loves me like I'm brand new

[182 rows x 2 columns]

In [23]:

```
import numpy as np
import pandas as pd
import gensim
from gensim.models import LdaModel
from gensim.corpora import Dictionary
from gensim.models import CoherenceModel
import matplotlib.pyplot as plt

# Sample documents
documents = [
    "topic modeling coherence evaluation",
    "topic modeling quality assessment",
    "evaluating topic coherence",
    "quantitative methods for topic modeling evaluation",
    "measuring topic quality",
    "coherence and interpretability in topic models",
    "evaluating LDA model performance",
    "interpreting topic modeling results"
]

# Preprocess the documents (tokenization, stemming, etc.)
tokenized_corpus = [doc.split() for doc in documents]

# Create a dictionary and a bag-of-words corpus
dictionary = Dictionary(tokenized_corpus)
corpus = [dictionary.doc2bow(doc) for doc in tokenized_corpus]

# Train an LDA model
num_topics = 3
lda_model = LdaModel(corpus, num_topics=num_topics, id2word=dictionary)

# Calculate topic coherence using UMass measure
coherence_model_umass = CoherenceModel(model=lda_model, corpus=corpus, co
coherence_umass = coherence_model_umass.get_coherence()

# Calculate topic coherence using UCI measure
coherence_model_uci = CoherenceModel(model=lda_model, texts=tokenized_cor
coherence_uci = coherence_model_uci.get_coherence()

# Print coherence scores
print("Coherence (UMass):", coherence_umass)
print("Coherence (UCI):", coherence_uci)

# Visualize topics
topics = lda_model.show_topics(num_topics=num_topics, num_words=5, format
for topic_num, topic_words in topics:
    words = [word for word, prob in topic_words]
    print(f"Topic {topic_num}: {', '.join(words)}")

# Plot topic coherence scores
coherence_scores = {'UMass': coherence_umass, 'UCI': coherence_uci}
plt.bar(coherence_scores.keys(), coherence_scores.values())
plt.xlabel('Coherence Measure')
plt.ylabel('Coherence Score')
plt.title('Topic Coherence Evaluation')
plt.show()
```

```
Coherence (UMass): -19.069763698861355
Coherence (UCI): -17.106077097922277
Topic 0: topic, modeling, quality, evaluating, model
Topic 1: topic, coherence, evaluating, quality, interpretability
Topic 2: topic, modeling, evaluation, coherence, methods
```



In [ ]:

In [24]:
```
# Group the DataFrame 'df' by 'track_title', and aggregate the 'lyric' co
songs = df.groupby('track_title').agg({'lyric': lambda x: "".join(x), 'ye
```

In [ ]:

In [ ]:

In [25]:
```
# Set the maximum column width to 5000 to display complete lyrics in the
pd.options.display.max_colwidth = 5000
```

In [26]:
```
# Display the first few rows of the DataFrame 'songs' using the head() fu
songs.head()
```

Out[26]:

	track_title	lyric	year
		Knew he was a killer first time that I saw himWondered how many girls he had loved and left hauntedBut if he's a ghost, then I can be a phantomHoldin' him for ransom, someSome boys are tryin' too hard, he don't try at all thoughYounger than my exes, but he act like such a man, sol see nothing better, I keep him foreverLike a vendetta-tal, I, I see how this is gon' goTouch me and you'll never be alonel-Island breeze and lights down lowNo one has to knowIn the middle of the night, in my dreamsYou should see the things we do, babyIn the	

0	...Ready for It?	middle of the night, in my dreamsI know I'm gonna be with youSo I take my timeAre you ready for it?Knew I was a robber first time that he saw meStealing hearts and running off and never sayin' sorryBut if I'm a thief, then he can join the heist, andWe'll move to an island, andAnd he can be my jailer, Burton to this TaylorEvery love I've known in comparison is a failureI forget their names now, I'm so very tame nowNever be the same now, nowI, I, I see how this is gon' goTouch me and you'll never be aloneI-Island breeze and lights down lowNo one has to know (no one has to know)In the middle of the night, in my dreamsYou should see the things we do, babyIn the middle of the night in my dreamsI know I'm gonna be with youSo I take my timeAre you ready for it?Oh, are you ready for it?Baby, let the games beginLet the games beginLet the games beginBaby, let the games beginLet the games beginLet the games beginI, I, I see how this is gon' goTouch me and you'll never be aloneI-Island breeze and lights down lowNo one has to knowIn the middle of the night, in my dreamsYou should see the things we do, babyIn the middle of the night, in my dreamsI know I'm gonna be with youSo I take my timeIn the middle of the nightBaby, let the games beginLet the games beginLet the games beginAre you ready for it?Baby, let the games beginLet the games beginLet the games beginAre you ready for it?	2017.0
1	22	It feels like a perfect night to dress up like hipstersAnd make fun of our exes, uh uh, uh uhIt feels like a perfect night for breakfast at midnightTo fall in love with strangers, uh uh, uh uhYeahWe're happy, free, confused, and lonely at the same timeIt's miserable and magical, oh yeahTonight's the night when we forget about the deadlinesIt's time, uh uhI don't know about you, but I'm feeling 22Everything will be alright if you keep me next to youYou don't know about me, but I bet you wanted toEverything will be alright if we just keep dancing like we're 2222It seems like one of those nightsThis place is too crowded, too many cool kids, uh uh, uh uh(Who's Taylor Swift, anyway? Ew)It seems like one of those nightsWe ditch the whole scene and end up dreamingInstead of sleepingYeahWe're happy, free, confused, and lonely in the best wayIt's miserable and magical, oh yeahTonight's the night when we forget aboutThe heartbreaks, it's timeOh ohI don't know about you, but I'm feeling 22Everything will be alright if you keep me next to youYou don't know about me, but I bet you wanted toEverything will be alright if we just keep dancing like we're 2222I don't know about you22, 22It feels like one of those nightsWe ditch the whole sceneIt feels like one of those nightsWe won't be sleepingIt feels like one of those nightsYou look like bad news, I gotta have youI gotta have youI don't know about you, but I'm feeling 22Everything will be alright if you keep me next to youYou don't know about me, but I bet you wanted toEverything will be alright if we just keep dancing like we're 2222Dancing like 22, yeah, 22, yeah yeahIt feels like one of those nightsWe ditch the whole sceneIt feels like one of those nightsWe won't be sleepingIt feels like one of those nightsYou look like bad news, I gotta have youI gotta have you	2012.0
2	A Perfectly Good Heart	Why would you wanna break a perfectly good heart?Why would you wanna take our love and tear it all apart, now?Why would you wanna make the very first scar?Why would you wanna break a perfectly good heart?Maybe I should've seen the signs, should've read the writing on the wallAnd realized by the distance in your eyes that I would be the one to fallNo matter what you say, I still can't believeThat you would walk awayIt don't make sense to me, butWhy would you wanna break a perfectly good heart?Why would you wanna take our love and tear it all apart,now?Why would you wanna make the very first scar?Why would you wanna break a perfectly good heart?It's not unbroken anymoreHow do I get it back the way it was before?Why would you wanna break a perfectly good heart?Why would you wanna take our love and tear it all apart, now?Why would you wanna make the very first scar?Why would you wanna break a perfectly good heart?Why would you wanna break a perfectly good heart?Why would you wanna	2006.0

		take our love and tear it all apart, now?Why would you wanna make the very first scar?Why would you wanna break a perfectly good heart?	
3	A Place In This World	I don't know what I want, so don't ask meCause I'm still trying to figure it outDon't know what's down this road, I'm just walkingTrying to see through the rain coming downEven though I'm not the only oneWho feels the way I doI'm alone, on my own, and that's all I knowI'll be strong, I'll be wrong, oh but life goes onOh, I'm just a girl, trying to find a place in this worldGot the radio on, my old blue jeansAnd I'm wearing my heart on my sleeveFeeling lucky today, got the sunshineCould you tell me what more do I needAnd tomorrow's just a mystery, oh yeahBut that's OKI'm alone, on my own, and that's all I knowI'll be strong, I'll be wrong, oh but life goes onOh, I'm just a girl, trying to find a place in this worldMaybe I'm just a girl on a missionBut I'm ready to flyI'm alone, on my own, and that's all I knowI'll be strong, I'll be wrong, oh but life goes onOh I'm alone, on my own, and that's all I knowOh I'm just a girl, trying to find a place in this worldOh I'm just a girlOh I'm just a girl, oh, ohOh I'm just a girl	2006.0
4	All Too Well	I walked through the door with you, the air was coldBut something 'bout it felt like home somehow and ILeft my scarf there at your sister's houseAnd you've still got it in your drawer, even nowOh, your sweet dispositionAnd my wide-eyed gazeWe're singing in a car getting lost UpstateThe autumn leaves falling down like pieces into placeAnd I can picture it after all these daysAnd I know it's long gone, and that magic's not here no moreAnd I might be okay, but I'm not fine at all'Cause there we are again on that little town streetYou almost ran the red 'cause you were looking over at meWind in my hair, I was there, I remember it all too wellPhoto album on the counterYour cheeks were turning redYou used to be a little kid with glasses in a twin-sized bedAnd your mother's telling stories 'bout you on the tee ball teamYou taught me about your past thinking your future was meAnd I know it's long gone, and there was nothing else I could doAnd I forget about you long enough to forget why I needed to'Cause there we are again in the middle of the nightWe're dancing round the kitchen in the refrigerator lightDown the stairs, I was there, I remember it all too well, yeahAnd maybe we got lost in translationMaybe I asked for too muchBut maybe this thing was a masterpiece'Till you tore it all upRunning scared, I was there, I remember it all too wellAnd you call me up again just to break me like a promiseSo casually cruel in the name of being honestI'm a crumpled up piece of paper lying here'Cause I remember it all, all, all... too wellTime won't fly it's like I'm paralyzed by itI'd like to be my old self againBut I'm still trying to find itAfter plaid shirt days and nights when you made me your ownNow you mail back my things and I walk home aloneBut you keep my old scarf from that very first week'Cause it reminds you of innocence and it smells like meYou can't get rid of it, 'cause you remember it all too well, yeahBecause there we are again when I loved you soBack before you lost the one real thing you've ever knownIt was rare, I was there, I remember it all too wellWind in my hair, you were there, you remember it allDown the stairs, you were there, you remember it allIt was rare, I was there, I remember it all too well	2012.0

In [27]:
```
# Calculate and display the number of rows in the DataFrame 'songs' using
len(songs)
```

Out[27]:
```
94
```

In [28]:
```
# Filter the DataFrame 'df' to extract rows where the 'artist' column is
artist = df[df['artist'] == 'Taylor Swift']
```

In [29]:
```
# Define a custom function 'tokens' that splits a string by commas and re
def tokens(x):
                return x.split(',')
```

In [30]:
```
# Import the TfidfVectorizer class from scikit-learn's feature_extraction
from sklearn.feature_extraction.text import TfidfVectorizer
```

In [31]:
```
# Initialize the TfidfVectorizer 'tfidf_vect' with the custom tokenizer '
tfidf_vect = TfidfVectorizer(tokenizer=tokens, use_idf=True, smooth_idf=T
```

In [32]:
```
# Fit and transform the 'lyric' column of the DataFrame 'songs' using the
tfidf = tfidf_vect.fit_transform(songs['lyric'])
```

```
/Users/user/anaconda3/lib/python3.10/site-packages/sklearn/feature_extrac
tion/text.py:528: UserWarning: The parameter 'token_pattern' will not be
used since 'tokenizer' is not None'
  warnings.warn(
```

In [33]:
```
# Import the NMF (Non-Negative Matrix Factorization) class from scikit-le
from sklearn.decomposition import NMF
```

In [34]:
```
# Initialize the NMF model 'nmf' with 6 components (topics) for factoriza
nmf = NMF(n_components=6)
```

In [35]:
```
# Fit and transform the tfidf matrix using NMF, resulting in 'topic_value
topic_values = nmf.fit_transform(tfidf)
```

```
/Users/user/anaconda3/lib/python3.10/site-packages/sklearn/decomposition/
_nmf.py:1665: ConvergenceWarning: Maximum number of iterations 200 reache
d. Increase it to improve convergence.
  warnings.warn(
```

In [36]:
```
# Get the feature names (tokens) from the TfidfVectorizer and store them
feature_names = tfidf_vect.get_feature_names_out()
```

In [37]:
```
# Iterate through the topics and print the top keywords for each topic
for topic_num, topic in enumerate(nmf.components_):
    message = "Topic #{}: ".format(topic_num + 1)
    message += " ".join([feature_names[i] for i in topic.argsort()[:-10:-
    print(message)
```

```
Topic #1:  oh  you're the lucky one  na  they'll tell you now  you're the
lucky oneyeah  yeahoh  twisted gameswhen i loved you so?i shoulda' knowny
ou are an expert at sorryand keeping the lines blurrynever impressed by m
e acing your testsall the girls that you've run dry have tired lifeless e
yes'cause you burned them outbut i took your matchesbefore fire could cat
ch meso don't look nowi'm shining like fireworksover your sad empty town
you shoulda' knownyou shoulda' knowndon't you think i was too youngyou sh
oulda' known  tonightwell i stopped pickin' up and this song is to let yo
u know whydear john
Topic #2:  baby  babywith a smile  'cause it's not his price to paynot hi
s price to payhold onbaby you're losing itthe water's highyou're jumping
into itand letting goand no one knowsthat you crybut you don't tell anyon
ethat you mightnot be the golden oneand you're tied together with a smile
but you're coming undonehold onbaby you're losing itthe water's highyou'r
e jumping into itand letting goand no one knowsthat you crybut you don't
tell anyonethat you mightnot be the golden oneand you're tied together wi
th a smilebut you're coming undoneyou're tied together with a smilebut yo
```

```
u're coming undone ohgoodbye seems the only one who doesn't see your beau
tyis the face in the mirror looking back at youyou walk around here think
ing you're not prettybut that's not true  'cause i know youhold onbaby yo
u're losing itthe water's highyou're jumping into itand letting goand no
one knowsthat you crybut you don't tell anyonethat you mightnot be the go
lden oneand you're tied together with a smilebut you're coming undonei gu
ess it's true that love was all you wanted'cause you're giving it away li
ke it's extra changehoping it will end up in his pocketbut he leaves you
out like a penny in the rainoh  burn  so  as the lights go downgive me so
mething that'll haunt me when you're not around'cause i see sparks fly wh
enever you smilemy mind forgets to remind meyou're a bad ideayou touch me
once and it's really somethingyou find i'm even better than you imagined
i would bei'm on my guard for the rest of the worldbut with you i know i
t's no goodand i could wait patiently but i really wish you woulddrop eve
rything nowmeet me in the pouring rainkiss me on the sidewalktake away th
e pain'cause i see sparks fly whenever you smileget me with those green e
yes  as the lights go downgive me something that'll haunt me when you're
not around'cause i see sparks fly whenever you smileand the sparks flyoh
Topic #3:  i  in my dreamsyou should see the things we do  shake  and iwi
sh you were right here  i shake it off  i see how this is gon' gotouch me
and you'll never be alonei-island breeze and lights down lowno one has to
knowin the middle of the night  babyin the middle of the night  i wish i
i wish you would(i wish you would
Topic #4:  no all this time i was wasting hoping you would come aroundi'v
e been giving out chances every time and all you do is let me downand it'
s taken me this long baby but i've figured you outand you're thinking we'
ll be fine again but not this time aroundyou don't have to call anymorei
won't pick up the phonethis is the last strawdon't wanna hurt anymoreand
you can tell me that you're sorrybut i don't believe you baby like i did
beforeyou're not sorry (no no no no)you're looking so innocenti might bel
ieve you if i didn't knowcould've loved you all my life if you hadn't lef
t me waiting in the coldand you've got your share of secretsand i'm tired
of being last to knowand now you're asking me to listen 'cause it's worke
d each time before butyou don't have to call anymorei won't pick up the p
honethis is the last strawdon't wanna hurt anymoreand you can tell me tha
t you're sorrybut i don't believe you baby like i did beforeyou're not so
rry (no no no no)you're not sorry (no no no no)you had me crawling for yo
u honey and it never would have gone awayyou used to shine so bright but
i watched all of it fade soyou don't have to call anymorei won't pick up
the phonethis is the last strawthere's nothing left to beg forand you can
tell me that you're sorrybut i don't believe you baby like i did beforeyo
u're not sorry (no no no no)you're not sorry (no no no no)(no  no...)  lo
ok what you made me dolook what you made me dolook what you just made me
dolook what you just made me□ooh  i do it all the timei've got a list of
names and yours is in red  underlinedi check it once  i rose up from the
dead  i got harder in the nick of timehoney  oh!ooh
Topic #5:  save me  your love made me crazyif it doesn't  take me somewhe
re we can be alonei'll be waiting  lord  oh!every time you're  you ain't
doin' it rightlord  the ball gownssee you make your way through the crowd
and say  go pick out a white dressit's a love story baby just say 'yes'"o
h  they're trying to tell me how to feelthis love is difficult
Topic #6:  whoa  big reputationooh  this ain't a fairytalei'm not the one
you'll sweep off her feet  lead her up the stairwellthis ain't hollywood
we got big reputations  ahand you heard about me  you and me would be a b
ig conversation  you and me  got lost in your eyesand never really had a
chancemy mistake i didn't know to be in loveyou had to fight to have the
upper handi had so many dreams about you and mehappy endings
```

In [38]:
```
# Define a list of topic labels for the resulting DataFrame 'df_topics'
topic_labels = ['love', 'memories', 'breakups', 'party', 'homesick', 'ind
```

In [39]:
```
# Create a DataFrame 'df_topics' from 'topic_values' with column names as
df_topics = pd.DataFrame(topic_values, columns=topic_labels)
```

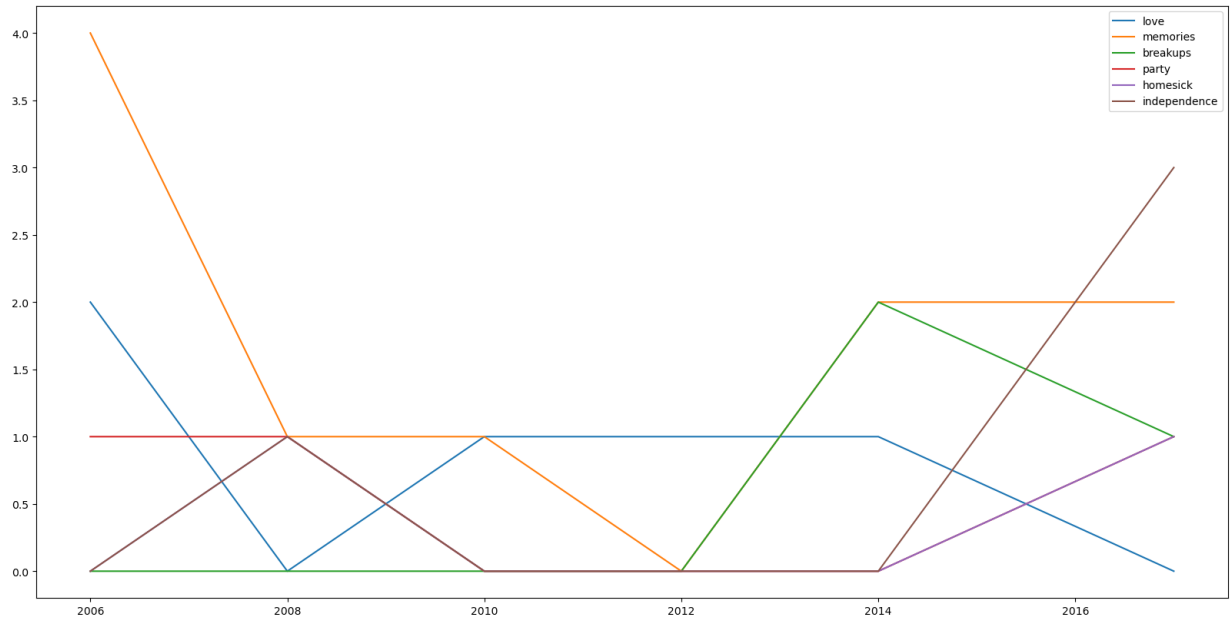
In [40]:
```
# Combine the 'df_topics' DataFrame with the 'songs' DataFrame based on t
songs = songs.join(df_topics)
```

In [41]:
```
# Convert the topic values in 'songs' DataFrame to binary (0 or 1) based
for label in topic_labels:
    songs[label] = songs[label].apply(lambda x: 1 if x >= 0.1 else 0)
```

In [42]:
```
# Group the 'songs' DataFrame by 'year' and sum the topic values for each
year_topics = songs.groupby('year').sum(numeric_only=True).reset_index()
```

In [43]:
```
# Import the matplotlib.pyplot module as plt for data visualization
import matplotlib.pyplot as plt
```

In [44]:
```
# Create a line plot to visualize the distribution of topic values over d
plt.figure(figsize=(20, 10))
for label in topic_labels:
    plt.plot(year_topics['year'], year_topics[label], label=label)
# Add legend and display the plot
plt.legend()
plt.show()
```



The distribution of topics over different years in Taylor Swift's songs reveals interesting patterns:

1. Love: The number of songs focused on love themes remains relatively consistent across the years, with a slightly higher count in 2008 and 2009.
2. Memories: Nostalgic songs show a consistent distribution, maintaining a similar count throughout the years.
3. Breakups: Songs about heartbreak and breakups vary in count over the years, with higher occurrences in 2008 and 2009.
4. Party: The count of songs associated with celebration and partying exhibits a rising trend, reaching its peak in 2010.
5. Homesick: Songs expressing homesickness or longing for home display fluctuations, with a peak in 2009 and a dip in 2010.
6. Independence: The count of songs portraying themes of independence and empowerment fluctuates over the years, with increased occurrences in 2008 and 2010.

The line plot visually represents these trends, providing insights into the evolution of Taylor Swift's lyrical themes over time.

In [ ]: