

%spark.pyspark

file\_path = 'hdfs://tw2770\_nyu\_edu/final-project/lyrics\_final.parquet'

keywords-selection

SPARK JOB FINISHED

```
df = spark.read.parquet(file_path)
df.show()
```

title	tag	artist	year	views	lyrics	id	rating
TO LOVE YOU.....	misc	Jhenny gonzaga	2015	0	I be glad if I tell...	2022005	1
Ur Stanton Nildab...	rock	Sam Asia	2015	0	We empower our cl...	2147320	1
Provident Kenwort...	misc	PropLadder	2015	0	Provident prestig...	2303729	1
Jest 1	misc	Tarlton's Jests	2015	0	To an ordinary in...	2312437	1
Tale of Genji Cha...	misc	Murasaki Shikibu	2022	0	CharacterDescript...	17564366	1
Coś o mnie	misc	Amadej Walczak	2022	0	Po ukończeniu stu...	17651058	1
Gerogan	rap	DJ Hamid Khareji	2021	0	In an interview, ...	16768076	1
Golem	rock	Akimbo (Band)	2003	0	For three years n...	16786967	1
Cavernous Black Eyes	rock	Akimbo (Band)	2003	0	All held back fea...	16786972	1
Skipping Stones	pop	Horace Pinker	2000	0	When i looked acr...	17705496	1
Carnival Nostalgia	pop	Horace Pinker	2000	0	I laughed with hi...	17705497	1
If You Fall	rock	Horace Pinker	2005	0	If you fall on yo...	17702041	1
Half Way	rock	Horace Pinker	2005	0	Hardly walking an...	17702042	1
Resonate So Real	rock	Horace Pinker	2005	0	He never walked a...	17702043	1
Belice	rock	Horace Pinker	2005	0	There's no need to	17702022	1

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:50 PM.

```
%spark.pyspark
df.printSchema()
```

FINISHED

```
root
 |-- title: string (nullable = true)
 |-- tag: string (nullable = true)
 |-- artist: string (nullable = true)
 |-- year: integer (nullable = true)
 |-- views: integer (nullable = true)
 |-- lyrics: string (nullable = true)
 |-- id: string (nullable = true)
 |-- rating: integer (nullable = true)
```

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:50 PM.

## Top Keywords

FINISHED

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:05:18 PM.

## Romove punctuation of lyrics column

FINISHED

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:05:54 PM.

```
%spark.pyspark
from pyspark.sql import SparkSession
```

FINISHED

```
from pyspark.ml.feature import Tokenizer, NGram
from pyspark.sql.functions import explode, col, lower, regexp_replace
```

```
# Remove punctuations and convert text to lowercase
```

## keywords-selection

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:50 PM.

```
%spark.pyspark
df.printSchema()
```

FINISHED

```
root
 |-- title: string (nullable = true)
 |-- tag: string (nullable = true)
 |-- artist: string (nullable = true)
 |-- year: integer (nullable = true)
 |-- views: integer (nullable = true)
 |-- lyrics: string (nullable = true)
 |-- id: string (nullable = true)
 |-- rating: integer (nullable = true)
```

Took 1 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:51 PM.

```
%spark. %%SPARK JOB (http://nyu-dataproc-sw-cfv5.c.hpc-dataproc-19b8.internal:46637/jobs/job?id=19) FINISHED
df.show()
```

title	tag	artist	year	views	lyrics	id	rating
TO LOVE YOU....	misc	Jhenny gonzaga	2015	0	be glad if i tell...	2022005	1
Ur Stanton Nildab...	rock	Sam Asia	2015	0	we empower our cl...	2147320	1
Provident Kenwort...	misc	PropLadder	2015	0	provident prestig...	2303729	1
Jest 1	misc	Tarlton's Jests	2015	0	to an ordinary in...	2312437	1
Tale of Genji Cha...	misc	Murasaki Shikibu	2022	0	characterdescript...	17564366	1
Coś o mnie	misc	Amadej Walczak	2022	0	po ukoczeniu stud...	17651058	1
Gerogan	rap	DJ Hamid Khareji	2021	0	in an interview h...	16768076	1
Golem	rock	Akimbo (Band)	2003	0	for three years n...	16786967	1
Cavernous Black Eyes	rock	Akimbo (Band)	2003	0	all held back fea...	16786972	1
Skipping Stones	pop	Horace Pinker	2000	0	when i looked acr...	17705496	1
Carnival Nostalgia	pop	Horace Pinker	2000	0	i laughed with hi...	17705497	1
If You Fall	rock	Horace Pinker	2005	0	if you fall on yo...	17702041	1
Half Way	rock	Horace Pinker	2005	0	hardly walking an...	17702042	1
Resonate So Real	rock	Horace Pinker	2005	0	he never walked a...	17702043	1

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:51 PM.

## Split lyrics column into tokens

FINISHED

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:51 PM.

```
%spark.pyspark
# Tokenize the lyrics into individual words
tokenizer = Tokenizer(inputCol="lyrics", outputCol="words")
words_data = tokenizer.transform(df).persist()
```

FINISHED

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:51 PM.



e", "we've", "you've", "they've", "isn't", "aren't", "wasn't", "weren't", "haven't", "has n't", "hadn't", "don't", "doesn't", "didn't", "won't", "wouldn't", "shan't", "shouldn't", "mustn't", "can't", "couldn't", 'cannot', 'could', "here's", "how's", "let's", 'ought', "th at's", "there's", "what's", "when's", "where's", "who's", "why's", 'would', ' ' , 'im', 'd

## keywords-selection

```
%spark.pyspark
print(extend_uni_stop_words)
```

FINISHED

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', 'your', 'yours', 'your
self', 'yourselves', 'he', 'him', 'his', 'himself', 'she', 'her', 'hers', 'herself', 'it',
'its', 'itself', 'they', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who',
'whom', 'this', 'that', 'these', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been',
'being', 'have', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'a
nd', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at', 'by', 'for', 'with',
'about', 'against', 'between', 'into', 'through', 'during', 'before', 'after', 'above', 'be
low', 'to', 'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under', 'again', 'furt
her', 'then', 'once', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both',
'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own',
'same', 'so', 'than', 'too', 'very', 's', 't', 'can', 'will', 'just', 'don', 'should', 'no
w', 'i'll', "you'll", "he'll", "she'll", "we'll", "they'll", 'i'd', "you'd", "he'd", "sh
e'd", "we'd", "they'd", 'i'm', "you're", "he's", "she's", "it's", "we're", "they're", 'i'v
e', "we've", "you've", "they've", "isn't", "aren't", "wasn't", "weren't", "haven't", "has
n't", "hadn't", "don't", "doesn't", "didn't", "won't", "wouldn't", "shan't", "shouldn't",
"mustn't", "can't", "couldn't", 'cannot', 'could', "here's", "how's", "let's", 'ought', "th
at's", "there's", "what's", "when's", "where's", "who's", "why's", 'would', ' ' , 'im', 'd
```

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:52 PM.

```
%spark.pyspark
# Remove stop words using the custom list
remover = StopWordsRemover(inputCol="words", outputCol="filtered_words")
remover.setStopWords(extend_uni_stop_words)
unigram_data = remover.transform(words_data)
```

FINISHED

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:52 PM.

```
%spark. SPARK JOB (http://nyu-dataproc-sw-cfv5.c.hpc-dataproc-19b8.internal:46637/jobs/job?id=21) FINISHED
unigram_data.show()
```

```
for Stanton mltad...|rockl      sum astut2015| 0|we empower our cl...|12147520| 1|
[we, empower, our...|empower, clients...|
|Provident Kenwort...|misc| PropLadder|2015| 0|provident prestig...|2303729| 1|
[provident, prest...|provident, prest...|
|      Jest 1|misc| Tarlton's Jests|2015| 0|to an ordinary in...|2312437| 1|
[to, an, ordinary...|ordinary, whitef...|
|Tale of Genji Cha...|misc|Murasaki Shikibu|2022| 0|characterdescript...|7564366| 1|
[characterdescrip...|characterdescrip...|
|      Coś o mniemisc| Amadej Walczak|2022| 0|po ukoczeniu stud...|7651058| 1|
[po, ukoczeniu, s...|po, ukoczeniu, s...|
|      Gerogan| rap|DJ Hamid Khareji|2021| 0|in an interview h...|6768076| 1|
[in, an, intervie...|interview, hamid...|
|      Golem|rockl Akimbo (Band)|2003| 0|for three years n...|6786967| 1|
[for, three, year...|three, years, st...|
|Cavernous Black Eyes|rockl Akimbo (Band)|2003| 0|all held back fea...|6786972| 1|
[all, held, back,...|held, fearing, d...|
|      Skipping Stones|popl Horace Pinker|2000| 0|when i looked acr...|7705496| 1|
[when, i, looked,...|looked, across, ...|
```

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:52 PM.

# keywords-selection

SPARK JOB FINISHED

```
%spark.pyspark
# Find all the unigram frequencies
unigrams_flattened = unigram_data.select(explode(col("filtered_words")).alias("unigram"))
unigram_counts = unigrams_flattened.groupBy("unigram").count()
unigram_counts_sorted = unigram_counts.orderBy(col("count").desc())
unigram_counts_sorted.show(100)

+-----+-----+
|      desc|count|
+-----+-----+
|      place|1684|
|      play|1683|
|      non|1674|
|      show|1661|
|      pain|1659|
|      che|1649|
|      lost|1603|
|      alone|1580|
|      thing|1573|
| without|1572|
|      words|1557|
|      hear|1551|
|      da|1550|
|      name|1544|
+-----+-----+
only showing top 100 rows
```

Took 3 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:55 PM.

```
%spark.pyspark
uni_words = ['like', 'life', 'love', 'yeah', 'shit', 'fuck', 'bitch', 'world', 'mind', 'he',
, 'stay', 'night', 'best']
```

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:55 PM.

## Bigram

FINISHED

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:06:26 PM.

```
%spark.pyspark
bigram_stop_words = ["", 'im', 'dont', 'know', 'got', 'get', 'time', 'one', 'cant', 'se',
'think', 'let', 'man', 'que', 'back', 'thats', 'feel', 'cause', 'still', 'day', 'away',
'interview', 'nigga', 'looks', 'look']
extend_bi_stop_words = default_stop_words + bigram_stop_words
```

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:55 PM.

```
%spark.pyspark
words_data.show()
```

# keywords-selection

```
[to, an, ordinary...|
|Tale of Genji Cha...|misc|Murasaki Shikibu|2022|    0|characterdescript...|7564366|    1|
|characterdescrip...|
|    Coś o mniemisc|    Amadej Walczak|2022|    0|po ukoczeniu stud...|7651058|    1|
|po, ukoczeniu, s...|
|    Gerogan| rap|DJ Hamid Khareji|2021|    0|in an interview h...|6768076|    1|
|in, an, interview...|
|    Golem|rock|    Akimbo (Band)|2003|    0|for three years n...|6786967|    1|
|for, three, year...|
|Cavernous Black Eyes|rock|    Akimbo (Band)|2003|    0|all held back fea...|6786972|    1|
|all, held, back,...|
|    Skipping Stones|pop|    Horace Pinker|2000|    0|when i looked acr...|7705496|    1|
|when i looked    |
```

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:55 PM.

```
%spark.pyspark
# Remove stop words using the custom list
remover = StopWordsRemover(inputCol="words", outputCol="filtered_words")
remover.setStopWords(extend_bi_stop_words)
bigrams_data = remover.transform(words_data)
```

FINISHED

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:55 PM.

```
%spark.FINISHED SPARK JOB (http://nyu-dataproc-sw-cfv5.c.hpc-dataproc-19b8.internal:46637/jobs/job?id=25) FINISHED
bigrams_data.show()
for Stanton inLadd...|rock|    Sam Astor|2015|    0|we empower our cl...|2147520|    1|
|we, empower, our...|empower, clients...| | | | | | |
|Provident Kenwort...|misc|    PropLadder|2015|    0|provident prestig...|2303729|    1|
|provident, prest...|provident, prest...|
|    Jest 1|misc|    Tarlton's Jests|2015|    0|to an ordinary in...|2312437|    1|
|to, an, ordinary...|ordinary, whitef...|
|Tale of Genji Cha...|misc|Murasaki Shikibu|2022|    0|characterdescript...|7564366|    1|
|characterdescrip...|characterdescrip...|
|    Coś o mniemisc|    Amadej Walczak|2022|    0|po ukoczeniu stud...|7651058|    1|
|po, ukoczeniu, s...|po, ukoczeniu, s...|
|    Gerogan| rap|DJ Hamid Khareji|2021|    0|in an interview h...|6768076|    1|
|in, an, interview...|hamid, khareji, ...|
|    Golem|rock|    Akimbo (Band)|2003|    0|for three years n...|6786967|    1|
|for, three, year...|three, years, st...|
|Cavernous Black Eyes|rock|    Akimbo (Band)|2003|    0|all held back fea...|6786972|    1|
|all, held, back,...|held, fearing, d...|
|    Skipping Stones|pop|    Horace Pinker|2000|    0|when i looked acr...|7705496|    1|
|when, i, looked,...|looked, across, ...|
```

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:55 PM.

```
%spark.pyspark
bigram = NGram(n=2, inputCol="filtered_words", outputCol="bigrams")
bigram_data = bigram.transform(bigrams_data)
```

FINISHED

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:56 PM.

```
%spark.FINISHED SPARK JOB (http://nyu-dataproc-sw-cfv5.c.hpc-dataproc-19b8.internal:46637/jobs/job?id=26) FINISHED
bigram_data.show()
```

## keywords-selection

```
[provident, prest...|provident, prest...|provident presti...|
|          Jest 1|misc|Tarlton's Jests|2015|    0|to an ordinary in...|2312437|    1|
[to, an, ordinary...|ordinary, whitef...|ordinary whitefr...|
|Tale of Genji Cha...|misc|Murasaki Shikibu|2022|    0|characterdescript...|7564366|    1|
|characterdescrip...|characterdescrip...|characterdescrip...|
|          Coś o mniemisc|Amadej Walczak|2022|    0|po ukoczeniu stud...|7651058|    1|
[po, ukoczeniu, s...|po, ukoczeniu, s...|po ukoczeniu, uk...|
|          Gerogan|rap|DJ Hamid Khareji|2021|    0|in an interview h...|6768076|    1|
[in, an, interview...|hamid, khareji, ...|hamid khareji, k...|
|          Golem|rock|Akimbo (Band)|2003|    0|for three years n...|6786967|    1|
[for, three, year...|three, years, st...|three years, yea...|
|Cavernous Black Eyes|rock|Akimbo (Band)|2003|    0|all held back fea...|6786972|    1|
[all, held, back,...|held, fearing, d...|held fearing, fe...|
|          Skipping Stones|pop|Horace Pinker|2000|    0|when i looked acr...|7705496|    1|
```

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:56 PM.

```
%spark.pyspark
# Explode the bigrams list to get individual bigrams as rows
bigrams_flattened = bigram_data.select(explode(col("bigrams")).alias("bigram"))
# Calculate frequency by grouping by bigram and counting the occurrences
bigram_counts = bigrams_flattened.groupBy("bigram").count()
# Sort the bigrams in descending order based on count
bigram_counts_sorted = bigram_counts.orderBy(col("count").desc())
```

FINISHED

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:02:56 PM.

```
%spark.pyspark
bigram_counts_sorted.show(100)
```

SPARK JOB FINISHED

```
|          want | 110|
|          non mil | 118|
|    leave alone| 118|
|          verse 2| 117|
|    suck dick | 117|
|    never stop | 117|
| never forget | 115|
|          y y | 114|
|          say | 113|
|    never never | 113|
|    years ago | 112|
|    dead girls | 111|
|    never go | 111|
|          want | 111|
|    left behind | 110|
```

+-----+-----+

only showing top 100 rows

Took 7 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:03:03 PM.

```
%spark.pyspark
bi_words = ['every night', 'need loving', 'nothing left', 'yo bitch', 'say goodbye', 'best
'say love', 'leave alone', 'never stop', 'never forget', 'years ago']
```

FINISHED

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:03:03 PM.

# Trigram

FINISHED

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:06:49 PM.

## keywords-selection

%spark.pyspark

FINISHED

```
trigram_stop_words = ["", 'im', 'dont', 'know', 'got', 'get', 'time', 'one', 'cant', 's',
    'think', 'let', 'man', 'que', 'back', 'thats', 'feel', 'cause', 'still', 'day', 'away',
    'interview', 'nigga', 'looks', 'look']
extend_tri_stop_words = default_stop_words + trigram_stop_words
```

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:03:03 PM.

%spark.pyspark

FINISHED

```
# Remove stop words using the custom list
remover = StopWordsRemover(inputCol="words", outputCol="filtered_words")
remover.setStopWords(extend_tri_stop_words)
trigrams_data = remover.transform(words_data)
```

Took 1 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:03:04 PM.

```
%spark.%%SPARK JOB (http://nyu-dataproc-sw-cfv5.c.hpc-dataproc-19b8.internal:46637/jobs/job?id=29) FINISHED
trigrams_data.show()
```

```
for Stanton mltad...|rockl      Sum Astut|2015| 0|we empower our cl...|2147520| 1|
[we, empower, our...|empower, clients...|
|Provident Kenwort...|misc|      PropLadder|2015| 0|provident prestig...|2303729| 1|
[provident, prest...|provident, prest...|
|      Jest 1|misc| Tarlton's Jests|2015| 0|to an ordinary in...|2312437| 1|
[to, an, ordinary...|ordinary, whitef...|
|Tale of Genji Cha...|misc|Murasaki Shikibu|2022| 0|characterdescript...|7564366| 1|
[characterdescrip...|characterdescrip...|
|      Coś o mniemisc| Amadej Walczak|2022| 0|po ukoczeniu stud...|7651058| 1|
[po, ukoczeniu, s...|po, ukoczeniu, s...|
|      Gerogan| rap|DJ Hamid Khareji|2021| 0|in an interview h...|6768076| 1|
[in, an, intervie...|hamid, khareji, ...|
|      Golem|rockl   Akimbo (Band)|2003| 0|for three years n...|6786967| 1|
[for, three, year...|three, years, st...|
|Cavernous Black Eyes|rockl   Akimbo (Band)|2003| 0|all held back fea...|6786972| 1|
[all, held, back,...|held, fearing, d...|
|      Skipping Stones| pop|   Horace Pinker|2000| 0|when i looked acr...|7705496| 1|
[when, i, looked,...|looked, across, ...|
```

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:03:04 PM.

%spark.pyspark

FINISHED

```
# Create trigrams using NGram
trigram = NGram(n=3, inputCol="filtered_words", outputCol="trigrams")
trigram_data = trigram.transform(trigrams_data)
```

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:03:04 PM.

```
%spark.%%SPARK JOB (http://nyu-dataproc-sw-cfv5.c.hpc-dataproc-19b8.internal:46637/jobs/job?id=30) FINISHED
trigram_data.show()
```



```

+-----+-----+-----+-----+-----+-----+-----+-----+
|          title|tag|          artist|year|views|          lyrics|          id|rating|
|words|filtered_words|          trigrams|

```

# keywords-selection

```

|    TO LOVE YOU....|misc|  Jhenny gonzaga|2015|    0|be glad if i tell...|2022005|    1|
[be, glad, if, i,...|[glad, tell, miss...|[glad tell miss, ...|
|Ur Stanton Nildab...|rock|    Sam Asia|2015|    0|we empower our cl...|2147320|    1|
[we, empower, our...|[empower, clients...|[empower clients ...|
|Provident Kenwort...|misc|    PropLadder|2015|    0|provident prestig...|2303729|    1|
[provident, prest...|[provident, prest...|[provident presti...|
|          Jest 1|misc| Tarlton's Jests|2015|    0|to an ordinary in...|2312437|    1|
[to, an, ordinary...|[ordinary, whitef...|[ordinary whitefr...|
|Tale of Genji Cha...|misc|Murasaki Shikibu|2022|    0|characterdescript...|7564366|    1|
[characterdescrip...|[characterdescrip...|[characterdescrip...|
|          Coś o mniemisc| Amadej Walczak|2022|    0|po ukoczeniu stud...|7651058|    1|

```

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:03:04 PM.

```

%spark.pyspark
# Explode the trigrams list to get individual trigrams as rows
trigrams_flattened = trigram_data.select(explode(col("trigrams")).alias("trigram"))
# Calculate frequency by grouping by trigram and counting the occurrences
trigram_counts = trigrams_flattened.groupBy("trigram").count()
# Sort the trigrams in descending order based on count
trigram_counts_sorted = trigram_counts.orderBy(col("count").desc())

```

FINISHED

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:03:04 PM.

```

%spark.pyspark
trigram_counts_sorted.show(100)

```

SPARK JOB FINISHED

```

|          hey hey hey|    44|
|          never|    44|
|    2021  2021|    43|
|          right |    43|
|          right|    42|
|          girl|    42|
|    2020|    42|
|    work da work|    40|
|          ha ha ha|    40|
|          da work da|    40|
|    peuttre auraitil|    39|
|          dans une fort|    39|
|    diabolique moins|    39|
|    renfermer un secret|    39|
|          rsonne le mal|    39|
|    tnbres locarina na|    39|
|          se rveille les|    39|

```

Took 8 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:03:12 PM.

```

%spark.pyspark
tri_words = ['juicy dead girls', 'ay ay ay', 'oh oh oh', 'na na na']

```

FINISHED

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:03:12 PM.

## Top Keywords

FINISHED

### keywords-selection

1. Emotion and Relationships: like, life, love, yeah, mind, heart, girl, die, baby, god, leave, best friend, say goodbye, really want, say love, leave alone, never stop, never forget
2. Hardship and Negativity: shit, fuck, bitch, world, pain, alone, nothing left, yo bitch, years ago
3. Lifestyle and Culture: money, stay, night, every night, need loving, new york, brand new, hip hop, juicy dead girls, ay ay ay, oh oh oh, na na na

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:13:13 PM.

```
%spark.pyspark
```

FINISHED

```
keywords = uni_words + bi_words + tri_words  
print(keywords)
```

```
['like', 'life', 'love', 'yeah', 'shit', 'fuck', 'bitch', 'world', 'mind', 'heart', 'girl',  
'die', 'money', 'baby', 'god', 'leave', 'best', 'alone', 'pain', 'stay', 'night', 'best',  
'every night', 'need loving', 'nothing left', 'yo bitch', 'say goodbye', 'best friend', 'ne  
ver ever', 'really want', 'new york', 'brand new', 'hip hop', 'say love', 'leave alone', 'n  
ever stop', 'never forget', 'years ago', 'juicy dead girls', 'ay ay ay', 'oh oh oh', 'na na  
na']
```

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:03:13 PM.

```
%spark.pyspark  
len(keywords)
```

FINISHED

```
42
```

Took 0 sec. Last updated by tw2770\_nyu\_edu at December 09 2024, 1:03:13 PM.

```
%spark.pyspark
```

READY