

▼ **BTS Word Cloud**

Importing Libraries

```
import pandas as pd
import re
import nltk
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
from wordcloud import WordCloud
import matplotlib.pyplot as plt
```

Importing the file

```
df = pd.read_csv('lyrics-v6.csv',encoding='latin1')
```

```
df.head()
```



	id	album_title	eng_album_title	album_rd	album_seq	track_title	kor_track_title	eng_track_title	lyrics	hidden_track	rem
0	0	2 Cool 4 Skool	2 Cool 4 Skool	2013-06-12	1	Intro: 2 Cool 4 Skool (ft. DJ Friz)	NaN	Intro: 2 Cool 4 Skool (ft. DJ Friz)	weâ□□re now going to progress to some steps\nw...	False	Fa
1	1	2 Cool 4 Skool	2 Cool 4 Skool	2013-06-12	2	We Are Bulletproof Pt.2	NaN	We Are Bulletproof Pt.2	(what) give it to me\n (what) be nervous\n (wh...	False	Fa
2	2	2 Cool 4 Skool	2 Cool 4 Skool	2013-06-12	3	Skit: Circle Room Talk	NaN	Skit: Circle Room Talk	rap monster: it was a big hit\nv: year 2006!\n...	False	Fa
3	3	2 Cool 4 Skool	2 Cool 4 Skool	2013-06-12	4	No More Dream	NaN	No More Dream	hey, what's your dream?\n hey, whatâ□□s your d...	False	Fa
4	4	2 Cool 4 Skool	2 Cool 4 Skool	2013-06-12	5	Interlude	NaN	Interlude	NaN	False	Fa

Next steps:

[Generate code with df](#)


 [View recommended plots](#)

[New interactive sheet](#)

Extracting Lyrics column

```
lyrics = df['lyrics']
```

```
lyrics.head()
```



	lyrics
0	weâ□□re now going to progress to some steps\nw...
1	(what) give it to me\n (what) be nervous\n (wh...
2	rap monster: it was a big hit\nv: year 2006!\n...
3	hey, what's your dream?\n hey, whatâ□□s your d...
4	NaN

```
# Removing Null Values
```

```
lyrics = lyrics.dropna()
```

```
lyrics.head()
```



lyrics

```
0  weâre now going to progress to some steps\nw...
1  (what) give it to me\n (what) be nervous\n (wh...
2  rap monster: it was a big hit\nv: year 2006!\n...
3  hey, what's your dream?\n hey, whatâs your d...
5  wanna be lovedâ\!\n\ndonâ wanna be fool wan...
```

```
# Removing special characters except a-zA-Z
```

```
lyrics = lyrics.apply(lambda x: re.sub(r'^a-zA-Z\s', '', x))
```

```
lyrics.head()
```



lyrics

```
0  were now going to progress to some steps\nwhic...
1  what give it to me\n what be nervous\n what th...
2  rap monster it was a big hit\nv year \nrap mon...
3  hey whats your dream\n hey whats your dream\n ...
5  wanna be loved\n\ndont wanna be fool wanna be ...
```

```
# Converting all the texts to lower case
```

```
lyrics = lyrics.apply(lambda x: x.lower())
```

```
# Tokenization
```

```
nltk.download('punkt')
```

```
df['lyrics_tokenized'] = lyrics.apply(lambda x: word_tokenize(x))
```



```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!
```

```
df.head()
```

	id	album_title	eng_album_title	album_rd	album_seq	track_title	kor_track_title	eng_track_title	lyrics	hidden_track	rem
0	0	2 Cool 4 Skool	2 Cool 4 Skool	2013-06-12	1	Intro: 2 Cool 4 Skool (ft. DJ Friz)	NaN	Intro: 2 Cool 4 Skool (ft. DJ Friz)	weâ□□re now going to progress to some steps\nw...	False	Fa
1	1	2 Cool 4 Skool	2 Cool 4 Skool	2013-06-12	2	We Are Bulletproof Pt.2	NaN	We Are Bulletproof Pt.2	(what) give it to me\n (what) be nervous\n (wh...	False	Fa
2	2	2 Cool 4 Skool	2 Cool 4 Skool	2013-06-12	3	Skit: Circle Room Talk	NaN	Skit: Circle Room Talk	rap monster: it was a big hit\nv: year 2006!\n...	False	Fa
3	3	2 Cool 4 Skool	2 Cool 4 Skool	2013-06-12	4	No More Dream	NaN	No More Dream	hey, what's your dream?\n hey, whatâ□□s your d...	False	Fa
4	4	2 Cool 4 Skool	2 Cool 4 Skool	2013-06-12	5	Interlude	NaN	Interlude	NaN	False	Fa

Next steps: [Generate code with df](#) [View recommended plots](#) [New interactive sheet](#)

```
# Drop all other columns from the dataframe except lyrics and lyrics_tokenized
df = df.drop(['id', 'album_title', 'eng_album_title', 'album_rd', 'album_seq','track_title','kor_track_title','eng_track_title','hidden_t

df.head()
```

	lyrics	lyrics_tokenized
0	weâ□□re now going to progress to some steps\nw...	[were, now, going, to, progress, to, some, ste...
1	(what) give it to me\n (what) be nervous\n (wh...	[what, give, it, to, me, what, be, nervous, wh...
2	rap monster: it was a big hit\nv: year 2006!\n...	[rap, monster, it, was, a, big, hit, v, year, ...
3	hey, what's your dream?\n hey, whatâ□□s your d...	[hey, whats, your, dream, hey, whats, your, dr...
4	NaN	NaN

Next steps: [Generate code with df](#) [View recommended plots](#) [New interactive sheet](#)

```
df=df.dropna()

df.head()
```

	lyrics	lyrics_tokenized
0	weâ□□re now going to progress to some steps\nw...	[were, now, going, to, progress, to, some, ste...
1	(what) give it to me\n (what) be nervous\n (wh...	[what, give, it, to, me, what, be, nervous, wh...
2	rap monster: it was a big hit\nv: year 2006!\n...	[rap, monster, it, was, a, big, hit, v, year, ...
3	hey, what's your dream?\n hey, whatâ□□s your d...	[hey, whats, your, dream, hey, whats, your, dr...
5	wanna be lovedâ□□\n\ndonâ□□t wanna be fool wan...	[wan. na. be. loved. dont. wan. na. be. fool. ...

Next steps: [Generate code with df](#) [View recommended plots](#) [New interactive sheet](#)

```
# Lemmatization

nltk.download('wordnet')
nltk.download('omw-1.4')
lemmatizer = WordNetLemmatizer()
df['lyrics_lemmatized'] = df['lyrics_tokenized'].apply(lambda x: [lemmatizer.lemmatize(word) for word in x])
```

```
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Package wordnet is already up-to-date!
[nltk_data] Downloading package omw-1.4 to /root/nltk_data...
[nltk_data] Package omw-1.4 is already up-to-date!
```

```
df.head()
```

	lyrics	lyrics_tokenized	lyrics_lemmatized
0	weâ□re now going to progress to some steps\nw...	[were, now, going, to, progress, to, some, ste...	[were, now, going, to, progress, to, some, ste...
1	(what) give it to me\n (what) be nervous\n (wh...	[what, give, it, to, me, what, be, nervous, wh...	[what, give, it, to, me, what, be, nervous, wh...
2	rap monster: it was a big hit\nv: year 2006!\n...	[rap, monster, it, was, a, big, hit, v, year, ...	[rap, monster, it, wa, a, big, hit, v, year, r...
3	hey, what's your dream?\n hey, whatâ□s your d...	[hey, whats, your, dream, hey, whats, your, dr...	[hey, whats, your, dream, hey, whats, your, dr...

Next steps: [Generate code with df](#) [View recommended plots](#) [New interactive sheet](#)

```
# Removing stopwords
```

```
nltk.download('stopwords')
stop_words = set(stopwords.words('english'))
df['lyrics_lemmatized_cleaned'] = df['lyrics_lemmatized'].apply(lambda x: [word for word in x if word not in stop_words])
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

```
df.head()
```

	lyrics	lyrics_tokenized	lyrics_lemmatized	lyrics_lemmatized_cleaned
0	weâ□re now going to progress to some steps\nw...	[were, now, going, to, progress, to, some, ste...	[were, now, going, to, progress, to, some, ste...	[going, progress, step, bit, difficult, ready...
1	(what) give it to me\n (what) be nervous\n (wh...	[what, give, it, to, me, what, be, nervous, wh...	[what, give, it, to, me, what, be, nervous, wh...	[give, nervous, one, end, bulletproof, bulletp...
2	rap monster: it was a big hit\nv: year 2006!\n...	[rap, monster, it, was, a, big, hit, v, year, ...	[rap, monster, it, wa, a, big, hit, v, year, r...	[rap, monster, wa, big, hit, v, year, rap, mon...
3	hey, what's your dream?\n hey, whatâ□s your d...	[hey, whats, your, dream, hey, whats, your, dr...	[hey, whats, your, dream, hey, whats, your, dr...	[hey, whats, dream, hey, whats, dream, hey, wh...
5	wanna be lovedâ□\n\ndonâ□t wanna be fool wan...	[wan, na, be, loved, dont, wan, na. be. fool. ...	[wan, na, be, loved, dont, wan, na. be. fool. ...	[wan, na, loved, dont, wan, na, fool, wan. na....

Next steps: [Generate code with df](#) [View recommended plots](#) [New interactive sheet](#)

```
# Creating word cloud
```

```
text = ' '.join(df['lyrics_lemmatized_cleaned'].apply(lambda x: ' '.join(x)))
wordcloud = WordCloud(width=800, height=800, background_color='white').generate(text)
```

```
plt.figure(figsize=(8, 8), facecolor=None)
plt.imshow(wordcloud)
plt.axis('off')
plt.tight_layout(pad=0)
plt.show()
```

