```
In [1]:  import pandas as pd
         import numpy as np
         import re
         df1 = pd.read_csv("comments.csv")
         df1.head()
```

Out[1]:

	Name	People_URL	Page_Or_Person	Post_U
0	Andreas Kenotaph	https://m.facebook.com/andreas.inkmann	Person	https://m.facebook.com/story.ph story_fbid=3
1	Top fan\nLuca Maccarone	https://m.facebook.com/luca.maccarone.7	Person	https://m.facebook.com/story.ph story_fbid=3
2	Lia Silva	https://m.facebook.com/alma.lilly.aho	Person	https://m.facebook.com/story.ph story_fbid=3
3	Author\nShankra Festival	https://m.facebook.com/shankrafestival/	Page	https://m.facebook.com/story.ph story_fbid=3
4	Author\nShankra Festival	https://m.facebook.com/shankrafestival/	Page	https://m.facebook.com/story.ph story_fbid=3

```
In [2]:  df1.shape
```

Out[2]: (741, 5)

```
In [3]:  df2 = df1.drop(['People_URL','Post_URL','Page_Or_Person','Name'], axis = 1)
         df2.head()
```

Out[3]:

	Comment
0	Felt good to give you something back for all t...
1	I already can't wait for next year ☺ It will b...
2	Is the line up remaining the same?▯
3	yes :)
4	Yes, tickets for the 2020's edition will grant...

```
In [4]:  df3 = df2.astype(str).apply(lambda x: x.str.encode('ascii', 'ignore').str.decode('ascii
         df3.head(5)
```

```
Out[4]:
```

	Comment
0	Felt good to give you something back for all t...
1	I already cant wait for next year It will be ...
2	Is the line up remaining the same?
3	yes :)
4	Yes, tickets for the 2020's edition will grant...

```
In [5]: spec_chars = ["!',$','%','^','*','`','~','!','_','-','{','\','':','';','"','','<',',',',>
         for char in spec_chars:
             df3['Comment'] = df3['Comment'].str.replace(char, ' ')
```

```
In [6]: df3.head(8)
```

```
Out[6]:
```

	Comment
0	Felt good to give you something back for all t...
1	I already cant wait for next year It will be ...
2	Is the line up remaining the same
3	yes :)
4	Yes tickets for the 2020's edition will grant...
5	Lucas Bourget meme line up ^^
6	Ela Httniegir
7	Aguardo com muita serenidade pela minha primei...

```
In [7]: df3['Comment'] = df3['Comment'].astype(str).str.replace('aaa', 'a')
```

```
In [8]: df3.head(5)
```

```
Out[8]:
```

	Comment
0	Felt good to give you something back for all t...
1	I already cant wait for next year It will be ...
2	Is the line up remaining the same
3	yes :)
4	Yes tickets for the 2020's edition will grant...

```
In [9]: df3.replace('', np.nan, inplace=True)
         df3.head(5)
```

```
Out[9]:
```

	Comment
0	Felt good to give you something back for all t...
1	I already cant wait for next year It will be ...

	Comment
2	Is the line up remaining the same
3	yes :)
4	Yes tickets for the 2020's edition will grant...

```
In [22]:  df4= df3.Comment.str.split(expand=True).stack()
          df4.loc[t.str.len() >= 4].groupby(level=0).apply(' '.join)
```

```
Out[22]:  0      Felt good give something back efforts last yea...
          1      already cant wait next year will even greater ...
          2                          line remaining same
          4      tickets 2020's edition will grant access next ...
          5                          Lucas Bourget meme line
                               ...
          733                          Yeah
          734              media1.tenor.co media1.tenor.co
          735                          yeeeeeees
          736                          Retro mega nice
          740                          Gaan
          Length: 687, dtype: object
```

```
In [ ]:  df4.to_csv('wordcloudfinal.csv', index=False)
```

```
In [10]:  from subprocess import check_output
          from wordcloud import WordCloud, STOPWORDS
```

```
In [11]:  #mpl.rcParams['figure.figsize']=(8.0,6.0)     #(6.0,4.0)
          mpl.rcParams['font.size']=12                  #10
          mpl.rcParams['savefig.dpi']=100               #72
          mpl.rcParams['figure.subplot.bottom']=.1
```

```
In [12]:  stopwords = set(STOPWORDS)
          data = pd.read_csv("wordcloudfinal.csv")
```

```
In [13]:  wordcloud = WordCloud(
                              background_color='white',
                              stopwords=stopwords,
                              max_words=400,
                              max_font_size=60,
                              random_state=72
                              ).generate(str(data['Comment']))
```

```
In [14]:  print(wordcloud)
          fig = plt.figure(1)
          plt.imshow(wordcloud)
          plt.axis('off')
          plt.show()
          fig.savefig("word1.png", dpi=900)
```

```
<wordcloud.wordcloud.WordCloud object at 0x094E5E08>
```