```python
import pandas as pd

tweets = pd.read_csv(r'C:\Users\daksh\Desktop\SMA Tweets\tweets.csv')

tweets
```

C:\Users\daksh\anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3165: DtypeWarning: Columns (0,1,2,3,4,5,6,7,11,12,13,14,15,16,17,18) have mixed types.Specify dtype option on import or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,

Out[1]:

| | Datetime | Tweet Id | Text | Username | |
|---|---|---|---|---|---|
| 0 | 2022-02-24 03:12:47+00:00 | 1496684505247141888.0 | UA Massive explosions rocks #Kharkiv. #Russia ... | IdeologyWars | https:// |
| 1 | 2022-02-24 03:18:54+00:00 | 1496686044275695616.0 | UA Kharkiv is officially being struck by major... | IdeologyWars | https:// |
| 2 | 2022-02-24 03:22:42+00:00 | 1496687000375726080.0 | UA More angles on that strike in Kharkiv. #Rus... | IdeologyWars | https:// |
| 3 | 2022-02-24 03:25:36+00:00 | 1496687731434565632.0 | UA BM-21 Grad strikes opening on #Mariupol cit... | IdeologyWars | https:// |
| 4 | 2022-02-24 03:27:28+00:00 | 1496688201242759168.0 | UA Damage caused by strike in Kharkiv... #Russ... | IdeologyWars | https:// |
| ... | ... | ... | ... | ... | |
| 399851 | 2022-03-27 23:59:26+00:00 | 1508232258037112834 | Russia's mass media and telecommunications wat... | RenKeNews | https://tw |
| 399852 | 2022-03-27 23:59:27+00:00 | 1508232263930105861 | @tony07653102 @CNEWS Pour comprendre un confli... | philmarso | https:// |
| 399853 | 2022-03-27 23:59:34+00:00 | 1508232290413219841 | Putin ukrajinske vojnike naziva neonacizmom, a... | dicke5 | https:// |
| 399854 | 2022-03-27 23:59:44+00:00 | 1508232332289126401 | Watch: amid war ruins in #Ukraine, unbowed cla... | vsdawson_MUFC | https://twi |
| 399855 | 2022-03-27 23:59:45+00:00 | 1508232337229852679 | #britishsummertime started on Sunday 27th.\n\n... | satillus | https |

399856 rows × 20 columns

In [10]:

```python
len(tweets)
```

Out[10]:

399856

In [2]:

```python
index = [ i for i in range(len(tweets))]
```
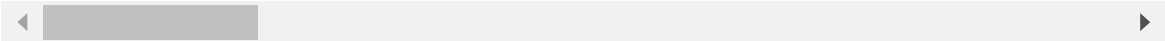
In [3]:

```python
tweets['Index'] = index
```

```
tweets.set_index('Index')
```

| Index | Datetime | Tweet Id | Text | Username | |
|---|---|---|---|---|---|
| 0 | 2022-02-24 03:12:47+00:00 | 1496684505247141888.0 | ᵤₐ Massive explosions rocks #Kharkiv. #Russia ... | IdeologyWars | https:// |
| 1 | 2022-02-24 03:18:54+00:00 | 1496686044275695616.0 | ᵤₐ Kharkiv is officially being struck by major... | IdeologyWars | https:// |
| 2 | 2022-02-24 03:22:42+00:00 | 1496687000375726080.0 | ᵤₐ More angles on that strike in Kharkiv. #Rus... | IdeologyWars | https:// |
| 3 | 2022-02-24 03:25:36+00:00 | 1496687731434565632.0 | ᵤₐ BM-21 Grad strikes opening on #Mariupol cit... | IdeologyWars | https:// |
| 4 | 2022-02-24 03:27:28+00:00 | 1496688201242759168.0 | ᵤₐ Damage caused by strike in Kharkiv... #Russ... | IdeologyWars | https:// |
| ... | ... | ... | ... | ... | |
| 399851 | 2022-03-27 23:59:26+00:00 | 1508232258037112834 | Russia's mass media and telecommunications wat... | RenKeNews | https://tw |
| 399852 | 2022-03-27 23:59:27+00:00 | 1508232263930105861 | @tony07653102 @CNEWS Pour comprendre un confli... | philmarso | https:// |
| 399853 | 2022-03-27 23:59:34+00:00 | 1508232290413219841 | Putin ukrajinske vojnike naziva neonacizmom, a... | dicke5 | https:// |
| 399854 | 2022-03-27 23:59:44+00:00 | 1508232332289126401 | Watch: amid war ruins in #Ukraine, unbowed cla... | vsdawson_MUFC | https://twi |
| 399855 | 2022-03-27 23:59:45+00:00 | 1508232337229852679 | #britishsummertime started on Sunday 27th.\n\n... | satillus | https |

399856 rows × 20 columns

```python
import nltk
nltk.download('stopwords')
list_stopwords = stopwords.words('english')
print(list_stopwords)
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "yo
u're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselv
es', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'hersel
f', 'it', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'the
mselves', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'thes
e', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'hav
e', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'th
e', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'a
t', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through',
'during', 'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down',
'in', 'out', 'on', 'off', 'over', 'under', 'again', 'further', 'then', 'on
ce', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both',
'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'no
t', 'only', 'own', 'same', 'so', 'than', 'too', 'very', 's', 't', 'can',
'will', 'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll',
'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'couldn', "couldn't",
'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'hasn', "hasn't",
'haven', "haven't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn',
"mustn't", 'needn', "needn't", 'shan', "shan't", 'shouldn', "shouldn't",
'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'wouldn', "would
n't"]
```

```python
list_stopwords.append('UA')
```

```python
print(list_stopwords)
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "yo
u're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselv
es', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', 'hers', 'hersel
f', 'it', "it's", 'its', 'itself', 'they', 'them', 'their', 'theirs', 'the
mselves', 'what', 'which', 'who', 'whom', 'this', 'that', "that'll", 'thes
e', 'those', 'am', 'is', 'are', 'was', 'were', 'be', 'been', 'being', 'hav
e', 'has', 'had', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'th
e', 'and', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'a
t', 'by', 'for', 'with', 'about', 'against', 'between', 'into', 'through',
'during', 'before', 'after', 'above', 'below', 'to', 'from', 'up', 'down',
'in', 'out', 'on', 'off', 'over', 'under', 'again', 'further', 'then', 'on
ce', 'here', 'there', 'when', 'where', 'why', 'how', 'all', 'any', 'both',
'each', 'few', 'more', 'most', 'other', 'some', 'such', 'no', 'nor', 'no
t', 'only', 'own', 'same', 'so', 'than', 'too', 'very', 's', 't', 'can',
'will', 'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll',
'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'couldn', "couldn't",
'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'hasn', "hasn't",
'haven', "haven't", 'isn', "isn't", 'ma', 'mightn', "mightn't", 'mustn',
"mustn't", 'needn', "needn't", 'shan', "shan't", 'shouldn', "shouldn't",
'wasn', "wasn't", 'weren', "weren't", 'won', "won't", 'wouldn', "would
n't", 'UA']
```

```python
tweets_compress = pd.DataFrame(columns = tweets.columns)
```

```python
import random
for i in range(3000):
    index = random.randint(0, 10000)
    if tweets['Language'][index] == 'en':
        rows = tweets.loc[[index]]
        tweets_compress = tweets_compress.append(rows, ignore_index=True)
```
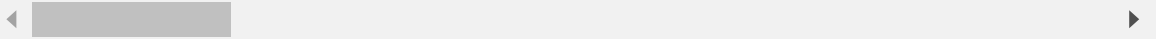
```
tweets_compress
```

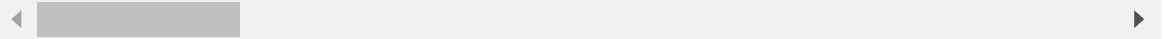| | Datetime | Tweet Id | Text | Username | |
|---|---|---|---|---|---|
| 0 | 2022-03-07 13:04:28+00:00 | 1500819674170286080.0 | I've just booked a two night stay in Kiev, Ukr... | chazsnell | https://twit |
| 1 | 2022-03-04 17:50:00+00:00 | 1499804364835364864.0 | Given #Russia's considerable foreign exchange ... | orfonline | https://twi |
| 2 | 2022-03-04 18:53:43+00:00 | 1499820399202295808.0 | Wow, one of my favorite actresses thank you mi... | AnaMelanyAmaya2 | https://twitter |
| 3 | 2022-03-20 22:01:08+00:00 | 1505665769824473088.0 | Ukraine's Zelenskiy presses Israel for missile... | sightmagazine | https://twit |
| 4 | 2022-03-08 22:32:50+00:00 | 1501325094324367360.0 | European oil receipts boosting Putin's war che... | MovieXen | https://twitte |
| ... | ... | ... | ... | ... | |
| 1854 | 2022-03-04 06:34:56+00:00 | 1499634479576621056.0 | A recap #PutinOligarchs #Oligarchy &amp; the #... | K08345975 | https://twitter |
| 1855 | 2022-03-16 04:19:43+00:00 | 1503949104518512640.0 | Anyone wants a good #daybyday #ukrainewar news... | driftlesstony | https://tv |
| 1856 | 2022-03-02 19:37:52+00:00 | 1499106733972729856.0 | Heerak Christian Kim, @GOP candidate for #USA ... | Heerak4Congress | https://twitte |
| 1857 | 2022-03-09 18:30:00+00:00 | 1501626371298238464.0 | RETREADCAST: Catch up on the latest #Retreadca... | RetreadBusiness | https://twit |
| 1858 | 2022-03-08 07:48:54+00:00 | 1501102643489697792.0 | Just 300 Ukrainian refugees granted visas to c... | MovieXen | https://twitte |

1859 rows × 21 columns

```
tweets_compress.sort_values('Index')
```

| | Datetime | Tweet Id | Text | Username | |
|---|---|---|---|---|---|
| 550 | 2022-02-24 03:36:10+00:00 | 1496690389364977664.0 | UA Air raid sirens are apparently in #Kyiv. #R... | IdeologyWars | https://twitter.c |
| 1128 | 2022-02-24 03:38:11+00:00 | 1496690894480752640.0 | UA Mariupol right now. #Russia #Ukraine #Ukrai... | IdeologyWars | https://twitter.c |
| 807 | 2022-02-24 03:43:16+00:00 | 1496692177107304448.0 | UA Saratov right now, turboprops can be heard.... | IdeologyWars | https://twitter.c |
| 1009 | 2022-02-24 03:46:01+00:00 | 1496692868089454592.0 | Ahahahaha @POTUS wont speak until tomorrow! Do... | ProudLatina213 | https://twitter.c |
| 1172 | 2022-02-24 03:54:07+00:00 | 1496694906517180416.0 | UA Dnepropetrovsk strike in the past few minut... | IdeologyWars | https://twitter.c |
| ... | ... | ... | ... | ... | |
| 637 | 2022-03-21 02:39:24+00:00 | 1505735797609226240.0 | "Go back home" #Ukraine people forcing 2 #Rus... | ali_saad_ali | https://twitter |
| 1062 | 2022-03-21 05:18:16+00:00 | 1505775779505836032.0 | Russia-Ukraine war LIVE: 4 dead in Kyiv shelli... | Oneindia | https://twitter.c |
| 1711 | 2022-03-21 05:33:03+00:00 | 1505779499912159232.0 | #JSW Steel Seshagiri Rao Speaks On Coking Coal... | mithun1500 | https://twitter.c |
| 1469 | 2022-03-21 05:48:59+00:00 | 1505783507867430912.0 | #UkraineConflict. Ammonia leak from Chemical p... | TQuantums | https://twitter.cor |
| 143 | 2022-03-21 06:08:12+00:00 | 1505788345724641280.0 | This is a tough read for any of us in our safe... | nick_carus | https://twitter.c |

1859 rows × 21 columns

```
index = [i for i in range(len(tweets_compress))]
```

```
tweets_compress['Index'] = index
```

```
tweets_compress.set_index('Index')
```

| | Datetime | Tweet Id | Text | Username | |
|---|---|---|---|---|---|
| **Index** | | | | | |
| **0** | 2022-03-07 13:04:28+00:00 | 1500819674170286080.0 | I've just booked a two night stay in Kiev, Ukr... | chazsnell | https://twi |
| **1** | 2022-03-04 17:50:00+00:00 | 1499804364835364864.0 | Given #Russia's considerable foreign exchange ... | orfonline | https://tw |
| **2** | 2022-03-04 18:53:43+00:00 | 1499820399202295808.0 | Wow, one of my favorite actresses thank you mi... | AnaMelanyAmaya2 | https://twitte |
| **3** | 2022-03-20 22:01:08+00:00 | 1505665769824473088.0 | Ukraine's Zelenskiy presses Israel for missile... | sightmagazine | https://twi |
| **4** | 2022-03-08 22:32:50+00:00 | 1501325094324367360.0 | European oil receipts boosting Putin's war che... | MovieXen | https://twitt |
| **...** | ... | ... | ... | ... | |
| **1854** | 2022-03-04 06:34:56+00:00 | 1499634479576621056.0 | A recap #PutinOligarchs #Oligarchy &amp; the #... | K08345975 | https://twitte |
| **1855** | 2022-03-16 04:19:43+00:00 | 1503949104518512640.0 | Anyone wants a good #daybyday #ukrainewar news... | driftlesstony | https:// |
| **1856** | 2022-03-02 19:37:52+00:00 | 1499106733972729856.0 | Heerak Christian Kim, @GOP candidate for #USA ... | Heerak4Congress | https://twitt |
| **1857** | 2022-03-09 18:30:00+00:00 | 1501626371298238464.0 | RETREADCAST: Catch up on the latest #Retreadca... | RetreadBusiness | https://twi |
| **1858** | 2022-03-08 07:48:54+00:00 | 1501102643489697792.0 | Just 300 Ukrainian refugees granted visas to c... | MovieXen | https://twitt |

1859 rows × 20 columns

In [81]:

```python
tweets_compress['Text Without Stop Words'] = ''
```

In [82]:

```python
temp = []
for text in tweets_compress['Text']:
    text = str(text)
    words = [word for word in text.split() if word.lower() not in list_stopwords]
    new_text = " ".join(words)
    temp.append(new_text)

tweets_compress['Text Without Stop Words'] = temp
```

In [87]:

```python
tweets_compress['Text Without Stop Words']
```

Out[87]:

```
0        I've booked two night stay Kiev, Ukraine airbn...
1        Given #Russia's considerable foreign exchange ...
2        Wow, one favorite actresses thank mila kunis c...
3        Ukraine's Zelenskiy presses Israel missile def...
4        European oil receipts boosting Putin's war che...
                               ...
1854     recap #PutinOligarchs #Oligarchy &amp; #GOP #P...
1855     Anyone wants good #daybyday #ukrainewar news m...
1856     Heerak Christian Kim, @GOP candidate #USA #Con...
1857     RETREADCAST: Catch latest #Retreadcast feature...
1858     300 Ukrainian refugees granted visas come UK ....
Name: Text Without Stop Words, Length: 1859, dtype: object
```

In [85]:

```python
tweets_compress.to_csv(r'C:\Users\daksh\Desktop\SMA Tweets\tweets_compress.csv')
```