

In [1]:

```
import json

# membaca data dalam format json
tweets_data = []
tweets_file = open('twitter_tweets22Mar.txt', "r")
for line in tweets_file:
    try:
        tweet = json.loads(line)
        tweets_data.append(tweet)
    except:
        continue
```

In [2]:

```
import pandas as pd

tweets = pd.DataFrame()
tweets["text"] = list(map(lambda tweet: tweet['text'], tweets_data))
tweets['lang'] = list(map(lambda tweet: tweet['lang'], tweets_data))
tweets['created_at'] = list(map(lambda tweet: tweet['created_at'], tweets_data))
tweets['country'] = list(map(lambda tweet: tweet['place']['country'] if tweet['place'] != None else None, tweets_data))

print ("jumlah data tweets :", len(tweets))
```

jumlah data tweets : 15805

In [3]:

```
# ##### Langkah kelima : memilih data yang berbahasa indonesia

tweetsIn = tweets[tweets.lang == 'in']
```

In [4]:

```
tweetsIn.head()
```

Out[4]:

	text	lang	created_at	country
4	W gabisa marah ama dia juga. Percuma marah ama...	in	Wed Mar 21 17:05:58 +0000 2018	Indonesia
5	mengaku sebagai istri di kehidupan yg sebelumn...	in	Wed Mar 21 17:06:01 +0000 2018	Indonesia
6	main game sambil tiduran 🤪 https://t.co/aSotVk...	in	Wed Mar 21 17:06:01 +0000 2018	Indonesia
8	Hari ini muka berasa bgt perih2 krn hrs di tus...	in	Wed Mar 21 17:06:24 +0000 2018	Indonesia
9	@anisaaurelias lah kok batre?	in	Wed Mar 21 17:06:34 +0000 2018	Indonesia

In [5]:

```
print ("jumlah data tweetsIn :", len(tweetsIn))
```

jumlah data tweetsIn : 11363

In [6]:

```
# ##### Langkah keenam : melakukan penyesuaian zona waktu
#
# konversi tipe data pada kolom 'created_at' menjadi format 'datetime'

# In[14]:
# import sys

# tweetsIn['created_at'] = pd.to_datetime(tweetsIn['created_at'], utc=True)

tweetsIn.index = pd.to_datetime(tweetsIn['created_at'], utc=True)
```

In []:

```
# merubah index tabel berdasarkan pada kolom 'created_at'

# In[15]:

# tweetsIn.index = tweetsIn.created_at
```

In [7]:

```
tweetsIn.head()
```

Out[7]:

	text	lang	created_at	country
created_at				
2018-03-21 17:05:58+00:00	W gabisa marah ama dia juga. Percuma marah ama...	in	Wed Mar 21 17:05:58 +0000 2018	Indonesia
2018-03-21 17:06:01+00:00	mengaku sebagai istri di kehidupan yg sebelumn...	in	Wed Mar 21 17:06:01 +0000 2018	Indonesia
2018-03-21 17:06:01+00:00	main game sambil tiduran 🤪 https://t.co/aSotVk...	in	Wed Mar 21 17:06:01 +0000 2018	Indonesia
2018-03-21 17:06:24+00:00	Hari ini muka berasa bgt perih2 krn hbs di tus...	in	Wed Mar 21 17:06:24 +0000 2018	Indonesia
2018-03-21 17:06:34+00:00	@anisaaurelias lah kok batre?	in	Wed Mar 21 17:06:34 +0000 2018	Indonesia

In [8]:

```
# melihat hasil perubahan index
```

```
# In[16]:
```

```
import pytz
from datetime import datetime
from pytz import timezone
```

```
JKT = pytz.timezone('Asia/Jakarta')
tweetsIn.index = tweetsIn.index.tz_convert(JKT)
```

In [9]:

```
# melihat hasil perubahan zona waktu di kolom index
```

```
# In[18]:
```

```
tweetsIn.head()
```

Out[9]:

	text	lang	created_at	country
created_at				
2018-03-22 00:05:58+07:00	W gabisa marah ama dia juga. Percuma marah ama...	in	Wed Mar 21 17:05:58 +0000 2018	Indonesia
2018-03-22 00:06:01+07:00	mengaku sebagai istri di kehidupan yg sebelumn...	in	Wed Mar 21 17:06:01 +0000 2018	Indonesia
2018-03-22 00:06:01+07:00	main game sambil tiduran 🤖 https://t.co/aSotVk...	in	Wed Mar 21 17:06:01 +0000 2018	Indonesia
2018-03-22 00:06:24+07:00	Hari ini muka berasa bgt perih2 krn hbs di tus...	in	Wed Mar 21 17:06:24 +0000 2018	Indonesia
2018-03-22 00:06:34+07:00	@anisaaurelias lah kok batre?	in	Wed Mar 21 17:06:34 +0000 2018	Indonesia

In [12]:

```
# indexer location
tweetsIn4 = tweetsIn.loc['2018-03-22']

tweetsIn4.head()
```

Out[12]:

	text	lang	created_at	country
created_at				
2018-03-22 00:05:58+07:00	W gabisa marah ama dia juga. Percuma marah ama...	in	Wed Mar 21 17:05:58 +0000 2018	Indonesia
2018-03-22 00:06:01+07:00	mengaku sebagai istri di kehidupan yg sebelumn...	in	Wed Mar 21 17:06:01 +0000 2018	Indonesia
2018-03-22 00:06:01+07:00	main game sambil tiduran 🤪 https://t.co/aSotVk...	in	Wed Mar 21 17:06:01 +0000 2018	Indonesia
2018-03-22 00:06:24+07:00	Hari ini muka berasa bgt perih2 krn hbs di tus...	in	Wed Mar 21 17:06:24 +0000 2018	Indonesia
2018-03-22 00:06:34+07:00	@anisaarelias lah kok batre?	in	Wed Mar 21 17:06:34 +0000 2018	Indonesia

In [13]:

```
tweetsIn4.tail()
```

Out[13]:

	text	lang	created_at	country
created_at				
2018-03-22 23:59:21+07:00	Martabak telur bersanding martabak manis, makn...	in	Thu Mar 22 16:59:21 +0000 2018	Indonesia
2018-03-22 23:59:26+07:00	Saya udah ikutan Uji Wawasan Redmi 5. Ikutan j...	in	Thu Mar 22 16:59:26 +0000 2018	Indonesia
2018-03-22 23:59:29+07:00	Bukan perasaannya yang salah.Waktunya aja belu...	in	Thu Mar 22 16:59:29 +0000 2018	Indonesia
2018-03-22 23:59:29+07:00	Katalog Produk Paket Bingkisan Lebaran 2018 @ ...	in	Thu Mar 22 16:59:29 +0000 2018	Indonesia
2018-03-22 23:59:55+07:00	Terima kasih kepada Bapak Duta Besar Belgia, Y...	in	Thu Mar 22 16:59:55 +0000 2018	Indonesia

In [14]:

```
print ("jumlah data tweets berbahasa Indonesia tanggal 1 April 2018 :", len(tweetsIn4))
```

```
df = pd.DataFrame({'tanggal' : ['1 April 2018'], # tanggal perlu diganti nih  
                  'jumlah tweets' : [len(tweetsIn4)]})
```

jumlah data tweets berbahasa Indonesia tanggal 1 April 2018 : 11332

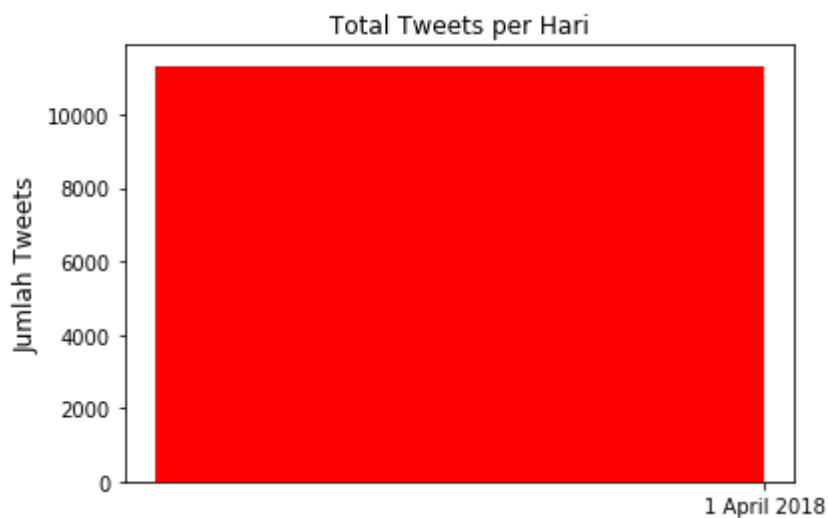
In [16]:

```
# melihat grafik bar total tweets per hari
```

```
# In[28]:
```

```
import matplotlib.pyplot as plt
```

```
fig, ax = plt.subplots()
x_pos = list(range(len(df)))
width = 2
plt.bar(x_pos, df['jumlah tweets'], width, alpha=1, color='r')
ax.set_ylabel('Jumlah Tweets', fontsize=12)
ax.set_title('Total Tweets per Hari', fontsize=12)
ax.set_xticks([p + 0.5 * width for p in x_pos])
ax.set_xticklabels(df['tanggal'])
plt.show()
```



In [17]:

```
# membuat tabel dan grafik data yang berisi jumlah total tweets per jam pada tanggal tertentu
```

```
import matplotlib.pyplot as plt
```

```
#import seaborn as sns
```

```
from pylab import *
```

```
get_ipython().magic('matplotlib inline') # %matplotlib inline magic command
```

```
plt.rcParams['figure.figsize'] = (15, 5)
```

In [18]:

```
# membuat tabel dan grafik data jumlah tweets pada tanggal <b>1 April 2018</b>
```

```
from pandas import Series
```

```
def f(x):
```

```
    return Series(dict(Number_of_tweets = x['text'].count()))
```

```
hourly_count25 = tweetsIn4.groupby(tweetsIn4.index.hour).apply(f)
```

```
print ("data selama %d jam pada tanggal 1 April 2018" % len(hourly_count25))
```

```
hourly_count25
```

data selama 24 jam pada tanggal 1 April 2018

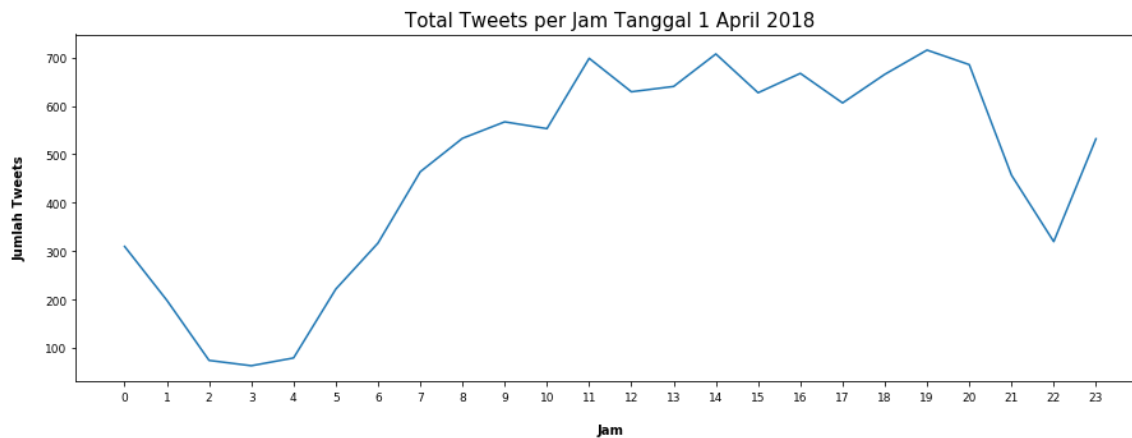
Out[18]:

	Number_of_tweets
created_at	
0	310
1	199
2	75
3	64
4	80
5	222
6	317
7	464
8	533
9	567
10	553
11	698
12	629
13	640
14	707
15	627
16	667
17	606
18	665
19	715
20	685
21	457
22	320
23	532

In [19]:

```
hourly_plot25 = hourly_count25['Number_of_tweets'].plot(kind='line')
hours = list(range(0,24))
xticks(np.arange(24), hours, rotation = 0, fontsize = 9)
hourly_plot25.set_title('Total Tweets per Jam Tanggal 1 April 2018', fontsize=15)
hourly_plot25.set_xlabel('Jam', weight='bold', labelpad=15)
hourly_plot25.set_ylabel('Jumlah Tweets', weight='bold', labelpad=15)
xticks(fontsize = 9, rotation = 0, ha= "center")
yticks(fontsize = 9)

hourly_plot25.tick_params(axis='x', pad=5)
```



In [20]:

```
hourly_plot25 = hourly_count25['Number_of_tweets'].plot(kind='bar')
hours = list(range(0,24))
xticks(np.arange(24), hours, rotation = 0, fontsize = 9)

hourly_plot25.set_title('Total Tweets per Jam Tanggal 1 April 2018', fontsize=15)
hourly_plot25.set_xlabel('Hour of the Day', weight='bold', labelpad=15)
hourly_plot25.set_ylabel('# Tweets (Messages)', weight='bold', labelpad=15)

xticks(fontsize = 9, rotation = 0, ha= "center")
yticks(fontsize = 9)
hourly_plot25.tick_params(axis='x', pad=5)
```

