

In [1]:

```
import json
```

```
# membaca data dalam format json
```

```
tweets_data = []
```

```
tweets_file = open('twitter_tweetsD2.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 : 18061

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
0	yg katanya besok pagi mau ikut race. tapi jam ...	in	Sat Mar 31 17:00:44 +0000 2018	Indonesia
1	@ikramarki @satriaoo Isinya bulu Ayam boiler y...	in	Sat Mar 31 17:00:46 +0000 2018	Indonesia
2	@an_zefa @detikcom Stju bro..wkwkwwk #anggurmerah	in	Sat Mar 31 17:00:47 +0000 2018	Indonesia
4	@SammiSoh Buat apa kita manusia tanggapin otak...	in	Sat Mar 31 17:01:01 +0000 2018	Indonesia
5	Bareng si cantik #BabyALUDRA .....duh hiburan ...	in	Sat Mar 31 17:01:05 +0000 2018	Indonesia

In [5]:

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

jumlah data tweetsIn : 12301

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 [7]:

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

# In[15]:

#tweetsIn.index = tweetsIn.created_at
```

In [8]:

```
tweetsIn.head()
```

Out[8]:

	text	lang	created_at	country
created_at				
2018-03-31 17:00:44+00:00	yg katanya besok pagi mau ikut race. tapi jam ...	in	Sat Mar 31 17:00:44 +0000 2018	Indonesia
2018-03-31 17:00:46+00:00	@ikramarki @satriaoo Isinya bulu Ayam boiler y...	in	Sat Mar 31 17:00:46 +0000 2018	Indonesia
2018-03-31 17:00:47+00:00	@an_zefa @detikcom Stju bro..wkwkww #anggurmerah	in	Sat Mar 31 17:00:47 +0000 2018	Indonesia
2018-03-31 17:01:01+00:00	@SammiSoh Buat apa kita manusia tanggapin otak...	in	Sat Mar 31 17:01:01 +0000 2018	Indonesia
2018-03-31 17:01:05+00:00	Bareng si cantik #BabyALUDRA .....duh hiburan ...	in	Sat Mar 31 17:01:05 +0000 2018	Indonesia

In [9]:

```
# 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 [10]:

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

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

```
tweetsIn.head()
```

Out[10]:

	text	lang	created_at	country
created_at				
2018-04-01 00:00:44+07:00	yg katanya besok pagi mau ikut race. tapi jam ...	in	Sat Mar 31 17:00:44 +0000 2018	Indonesia
2018-04-01 00:00:46+07:00	@ikramarki @satriaoo Isinya bulu Ayam boiler y...	in	Sat Mar 31 17:00:46 +0000 2018	Indonesia
2018-04-01 00:00:47+07:00	@an_zefa @detikcom Stju bro..wkwwkwk #anggurmerah	in	Sat Mar 31 17:00:47 +0000 2018	Indonesia
2018-04-01 00:01:01+07:00	@SammiSoh Buat apa kita manusia tanggapin otak...	in	Sat Mar 31 17:01:01 +0000 2018	Indonesia
2018-04-01 00:01:05+07:00	Bareng si cantik #BabyALUDRA .....duh hiburan ...	in	Sat Mar 31 17:01:05 +0000 2018	Indonesia

In [11]:

```
# indexer location
tweetsIn4 = tweetsIn.loc['2018-04-01']

tweetsIn4.head()
```

Out[11]:

	text	lang	created_at	country
created_at				
2018-04-01 00:00:44+07:00	yg katanya besok pagi mau ikut race. tapi jam ...	in	Sat Mar 31 17:00:44 +0000 2018	Indonesia
2018-04-01 00:00:46+07:00	@ikramarki @satriaoo Isinya bulu Ayam boiler y...	in	Sat Mar 31 17:00:46 +0000 2018	Indonesia
2018-04-01 00:00:47+07:00	@an_zefa @detikcom Stju bro..wkwkww #anggurmerah	in	Sat Mar 31 17:00:47 +0000 2018	Indonesia
2018-04-01 00:01:01+07:00	@SammiSoh Buat apa kita manusia tanggapin otak...	in	Sat Mar 31 17:01:01 +0000 2018	Indonesia
2018-04-01 00:01:05+07:00	Bareng si cantik #BabyALUDRA .....duh hiburan ...	in	Sat Mar 31 17:01:05 +0000 2018	Indonesia

In [12]:

```
tweetsIn4.tail()
```

Out[12]:

	text	lang	created_at	country
created_at				
<b>2018-04-01 23:59:22+07:00</b>	Duatahun lagi duaribu duapuluh <a href="https://t.co/ZF...">https://t.co/ZF...</a>	in	Sun Apr 01 16:59:22 +0000 2018	Indonesia
<b>2018-04-01 23:59:31+07:00</b>	Mohon maaf, nih. Kenapa makin lama makin berat...	in	Sun Apr 01 16:59:31 +0000 2018	Indonesia
<b>2018-04-01 23:59:50+07:00</b>	2 hari kemaren di Karawang, full kena panas mu...	in	Sun Apr 01 16:59:50 +0000 2018	Indonesia
<b>2018-04-01 23:59:51+07:00</b>	Waalaikumsalam <a href="https://t.co/XaJSFtXIUn">https://t.co/XaJSFtXIUn</a>	in	Sun Apr 01 16:59:51 +0000 2018	Indonesia
<b>2018-04-01 23:59:54+07:00</b>	Untuk keluarga besar BeloveTTM d'mna pun berad...	in	Sun Apr 01 16:59:54 +0000 2018	Indonesia

In [13]:

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

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

jumlah data tweets berbahasa Indonesia tanggal 1 April 2018 : 12296

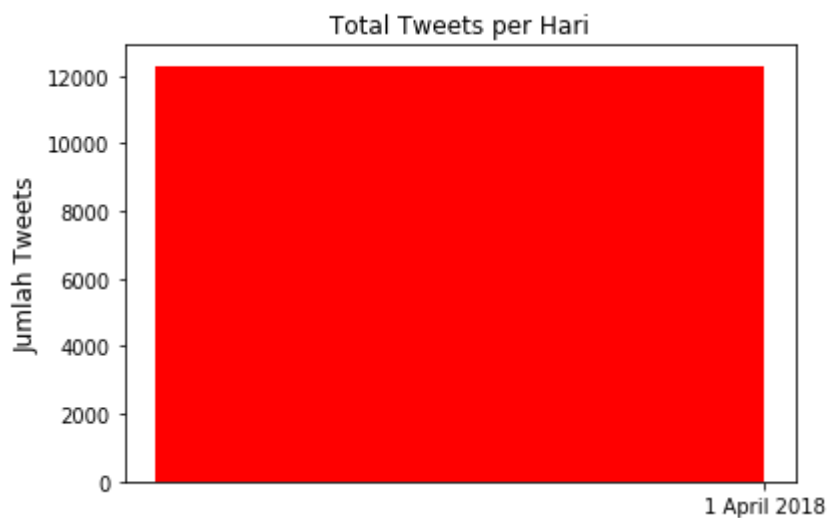
In [15]:

```
# 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 [16]:

```
# 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 [17]:

```
# 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[17]:

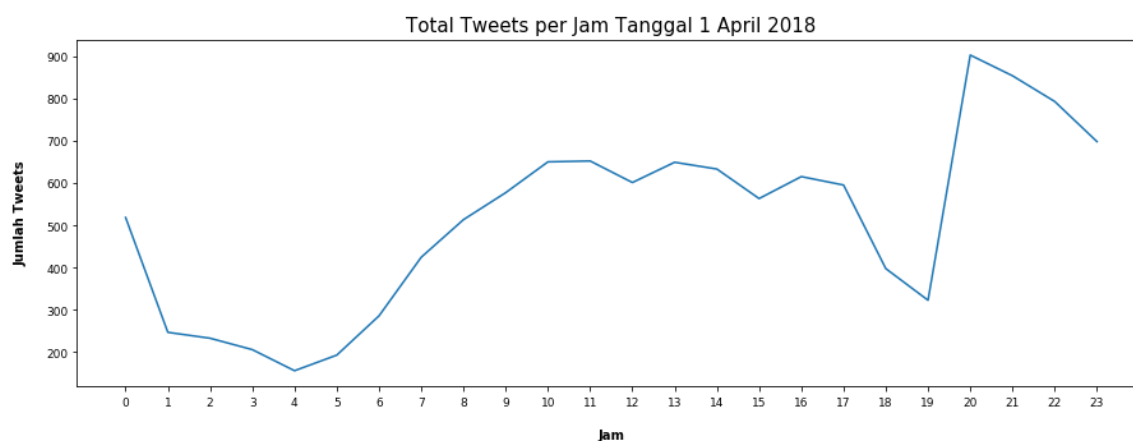
	Number_of_tweets
created_at	
0	519
1	247
2	233
3	206
4	156
5	193
6	286
7	425
8	514
9	578
10	651
11	653
12	602
13	650
14	634
15	564
16	616
17	596
18	398
19	323
20	904
21	855
22	794
23	699



In [18]:

```
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 [19]:

```
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)
```

