Workshop 8

```
from datetime import datetime
import pandas as pd
import matplotlib.pyplot as pyplot
```

Consider the following data points:

date	tick_numbers
2016-05-01 10:23:05.069722	3213
2016-05-01 10:23:05.119994	4324
2016-05-02 10:23:05.178768	2132
2016-05-02 10:23:05.230071	43242
2016-05-02 10:23:05.230071	4234
2016-05-02 10:23:05.280592	4234
2016-05-03 10:23:05.332662	4324
2016-05-03 10:23:05.385109	1245
2016-05-04 10:23:05.436523	1555
2016-05-04 10:23:05.486877	543345

Create a dataframe 'ts'

```
ts=
```

```
print ts
```

	date	tick_numbers
0	2016-05-01 10:23:05.069722	3213
1	2016-05-01 10:23:05.119994	4324
2	2016-05-02 10:23:05.178768	2132
3	2016-05-02 10:23:05.230071	43242
4	2016-05-02 10:23:05.230071	4234
5	2016-05-02 10:23:05.280592	4234
6	2016-05-03 10:23:05.332662	4324
7	2016-05-03 10:23:05.385109	1245
8	2016-05-04 10:23:05.436523	1555
9	2016-05-04 10:23:05.486877	543345

Convert ts['date'] from string to datetime. You can use ts.index.

```
ts.index=
```

Delete useless column with the command del 用del命令删除无用的列

```
del
```

```
print ts
```

```
In [17]: print ts
            tick_numbers
date
2016-05-01 10:23:05.069722      3213
2016-05-01 10:23:05.119994      4324
2016-05-02 10:23:05.178768      2132
2016-05-02 10:23:05.230071     43242
2016-05-02 10:23:05.230071      4234
2016-05-02 10:23:05.280592      4234
2016-05-03 10:23:05.332662      4324
2016-05-03 10:23:05.385109      1245
2016-05-04 10:23:05.436523      1555
2016-05-04 10:23:05.486877    543345
```

Print all data from 2016

Print all data from May 2016

Data after May 3rd, 2016

Remove all the data after May 2nd, 2016 using truncate

Count the number of data per timestamp

Mean value of ticks per day. You will use resample with a period of D and a method of mean.

Total value ticks per day. You will use sum and a period of D

Plot of the total of ticks per day

Create another dataframe

```
 np.random.seed(12345)
# create a dictionary
# df['ARCA'] = store np.random.randint(low=20000, high=30000, size=62)
# df['BARX'] = store np.random.randint(low=20000, high=30000, size=62)
# index = pd.date_range('4/1/2012', '6/1/2012')
# create the dataframe with the 3 components above

Print (df)
```

```
pd.DataFrame(volume,index=index).head()
Out[90]:
      ARCA  BARX
2012-04-01 24578 28633
2012-04-02 22177 26542
2012-04-03 23492 26554
2012-04-04 24094 21707
2012-04-05 24478 25568
```

Truncate the dataframe to get data (before='2012-04-04',after='2012-05-24')

Change the offset of the dataframe by pd.DateOffset(months=1, days=1)

Shift the dataframe by 1 day

Lag a variable 1 day

Aggregate into 2W-SUN (bi-weekly starting by Sunday) by summing up the value of each daily volumw

Aggregate into weeks by averaging up the value of each daily volume