pandas中的时间序列

python中datetime的复习
date_range
set_index
resample

time datetime timedelta复习

timedelta表示时间间隔

1. time

In [1]:
```
import time
```

In [2]:
```
time.time()        # 获得现在距离1970年初的时间戳
```

Out[2]:
```
1534068632.170157
```

In [3]:
```
time.localtime(1534066261)            # 将时间戳转为易读的时间
```

Out[3]:
```
time.struct_time(tm_year=2018, tm_mon=8, tm_mday=12, tm_hour=17, t
m_min=31, tm_sec=1, tm_wday=6, tm_yday=224, tm_isdst=0)
```

In [4]:
```
time.localtime(time.time())
```

Out[4]:
```
time.struct_time(tm_year=2018, tm_mon=8, tm_mday=12, tm_hour=18, t
m_min=10, tm_sec=32, tm_wday=6, tm_yday=224, tm_isdst=0)
```

In [5]:
```
time.strftime('%Y-%m-%d %H:%M:%S', time.localtime(time.time()))
# strftime() 可以将时间转化为指定的格式
```

Out[5]:
```
'2018-08-12 18:10:32'
```

In [6]:
```
time.strftime('%Y-%m-%d', time.localtime(time.time()))
```

Out[6]:
```
'2018-08-12'
```

In [7]:
```
time.strftime('%H:%M:%S', time.localtime(time.time()))
```

Out[7]:
```
'18:10:32'
```

In [8]:
```
local = time.strptime('2018-08-12 17:34:04', '%Y-%m-%d %H:%M:%S')
local
```

Out[8]:
```
time.struct_time(tm_year=2018, tm_mon=8, tm_mday=12, tm_hour=17, tm_min=34, tm_sec=4, tm_wday=6, tm_yday=224, tm_isdst=-1)
```

In [9]:
```
time.mktime(local)
```

Out[9]:
```
1534066444.0
```

In [10]:
```
local2 = time.strptime('2018-08-12 17', '%Y-%m-%d %H')
local2
```

Out[10]:
```
time.struct_time(tm_year=2018, tm_mon=8, tm_mday=12, tm_hour=17, tm_min=0, tm_sec=0, tm_wday=6, tm_yday=224, tm_isdst=-1)
```

In [11]:
```
time.mktime(local2)
```

Out[11]:
```
1534064400.0
```

2. datetime

In [12]:
```
from datetime import datetime
```

In [13]:
```
now = datetime.now()
now
```

Out[13]:
```
datetime.datetime(2018, 8, 12, 18, 10, 32, 314034)
```

In [14]:
```
now.year, now.month, now.day
```

Out[14]:
```
(2018, 8, 12)
```

In [15]:
```
now.strftime('%Y-%m-%d %H:%M:%S')
```

Out[15]:
```
'2018-08-12 18:10:32'
```

In [16]:
```
now.strftime('%Y-%m-%d')
```

Out[16]:
```
'2018-08-12'
```

In [17]:
```
now.strftime('%H:%M:%S')
```

Out[17]:
```
'18:10:32'
```

In [18]:
```
datetime.strptime('2018-08-12 17:39:50', '%Y-%m-%d %H:%M:%S')
```

Out[18]: datetime.datetime(2018, 8, 12, 17, 39, 50)

In [19]:
```
datetime(2018, 8, 12)
```

Out[19]: datetime.datetime(2018, 8, 12, 0, 0)

In [20]:
```
now.timestamp()          # 转化为时间戳
```

Out[20]: 1534068632.314034

In [21]:
```
datetime.fromtimestamp(1534066790.32124)
```

Out[21]: datetime.datetime(2018, 8, 12, 17, 39, 50, 321240)

3. timedelta

In [22]:
```
delta = datetime.now() - datetime(2017, 4, 25)
delta
```

Out[22]: datetime.timedelta(days=474, seconds=65432, microseconds=402076)

In [23]:
```
delta.days
```

Out[23]: 474

In [24]:
```
delta.seconds
```

Out[24]: 65432

In [25]:
```
delta.microseconds
```

Out[25]: 402076

date_range

In [26]:
```
import pandas as pd
import numpy as np
```

In [27]:
```
pd.date_range('2018-08-12', '2018-09-25')
```

Out[27]:
```
DatetimeIndex(['2018-08-12', '2018-08-13', '2018-08-14', '2018-08-15',
               '2018-08-16', '2018-08-17', '2018-08-18', '2018-08-19',
               '2018-08-20', '2018-08-21', '2018-08-22', '2018-08-23',
               '2018-08-24', '2018-08-25', '2018-08-26', '2018-08-27',
               '2018-08-28', '2018-08-29', '2018-08-30', '2018-08-31',
               '2018-09-01', '2018-09-02', '2018-09-03', '2018-09-04',
               '2018-09-05', '2018-09-06', '2018-09-07', '2018-09-08',
               '2018-09-09', '2018-09-10', '2018-09-11', '2018-09-12',
               '2018-09-13', '2018-09-14', '2018-09-15', '2018-09-16',
               '2018-09-17', '2018-09-18', '2018-09-19', '2018-09-20',
               '2018-09-21', '2018-09-22', '2018-09-23', '2018-09-24',
               '2018-09-25'],
              dtype='datetime64[ns]', freq='D')
```

In [28]:
```
pd.date_range('2018-08-12', '2018-09-25', freq='W')
```

Out[28]:
```
DatetimeIndex(['2018-08-12', '2018-08-19', '2018-08-26', '2018-09-02',
               '2018-09-09', '2018-09-16', '2018-09-23'],
              dtype='datetime64[ns]', freq='W-SUN')
```

In [29]:
```
pd.date_range('2018-08-12', '2018-09-25', freq='M')
```

Out[29]:
```
DatetimeIndex(['2018-08-31'], dtype='datetime64[ns]', freq='M')
```

In [30]:
```
pd.date_range('2018-08-12', '2018-09-25', freq='Q')         # D-天 W-周  M-月  Q-季度  A-年  H-小时  T-分  S-秒
```

Out[30]:
```
DatetimeIndex([], dtype='datetime64[ns]', freq='Q-DEC')
```

In [31]:
```
pd.date_range('2018-08-12', freq='W', periods=10)
```

Out[31]:
```
DatetimeIndex(['2018-08-12', '2018-08-19', '2018-08-26', '2018-09-02',
               '2018-09-09', '2018-09-16', '2018-09-23', '2018-09-30',
               '2018-10-07', '2018-10-14'],
              dtype='datetime64[ns]', freq='W-SUN')
```

In [32]:
```
data = {
    'time': pd.date_range('2018-08-12', freq='T', periods=200000),
    'cpu': np.random.randn(200000) + 10
}
df = pd.DataFrame(data, columns=['time', 'cpu'])
```

In [33]:
```
df.head()
```

Out[33]:

	time	cpu
0	2018-08-12 00:00:00	9.773730
1	2018-08-12 00:01:00	8.769905
2	2018-08-12 00:02:00	8.974288
3	2018-08-12 00:03:00	10.628897
4	2018-08-12 00:04:00	9.055220

In [34]:
```
df.tail()
```

Out[34]:

	time	cpu
199995	2018-12-28 21:15:00	9.421030
199996	2018-12-28 21:16:00	8.749773
199997	2018-12-28 21:17:00	10.713869
199998	2018-12-28 21:18:00	11.013194
199999	2018-12-28 21:19:00	10.920783

In [35]:
```
df[(df.time>='2018-08-12 08:00:00') & (df.time<='2018-08-12 08:10:00')]
```

Out[35]:

	time	cpu
480	2018-08-12 08:00:00	12.098545
481	2018-08-12 08:01:00	9.738498
482	2018-08-12 08:02:00	10.386033
483	2018-08-12 08:03:00	9.068647
484	2018-08-12 08:04:00	11.204582
485	2018-08-12 08:05:00	9.076552
486	2018-08-12 08:06:00	9.767191
487	2018-08-12 08:07:00	9.117590
488	2018-08-12 08:08:00	9.202167
489	2018-08-12 08:09:00	10.829297
490	2018-08-12 08:10:00	9.940771

set_index

In [36]:
```
df2 = df.set_index("time")
df2.head()
```

Out[36]:

	cpu
time	
2018-08-12 00:00:00	9.773730
2018-08-12 00:01:00	8.769905
2018-08-12 00:02:00	8.974288
2018-08-12 00:03:00	10.628897
2018-08-12 00:04:00	9.055220

set_index()等效于下面的代码：

In [37]:
```
s = pd.to_datetime(df.time)
s.head()
df.index = s
```

In [38]:
```
df.head()
```

Out[38]:

	time	cpu
time		
2018-08-12 00:00:00	2018-08-12 00:00:00	9.773730
2018-08-12 00:01:00	2018-08-12 00:01:00	8.769905
2018-08-12 00:02:00	2018-08-12 00:02:00	8.974288
2018-08-12 00:03:00	2018-08-12 00:03:00	10.628897
2018-08-12 00:04:00	2018-08-12 00:04:00	9.055220

In [39]:
```
df = df.drop('time', axis=1)
df.head()
```

Out[39]:

	cpu
time	
2018-08-12 00:00:00	9.773730
2018-08-12 00:01:00	8.769905
2018-08-12 00:02:00	8.974288
2018-08-12 00:03:00	10.628897
2018-08-12 00:04:00	9.055220

In [40]:
```
df['2018-08-12 08:00:00':'2018-08-12 08:10:00']
```

Out[40]:

	cpu
time	
2018-08-12 08:00:00	12.098545
2018-08-12 08:01:00	9.738498
2018-08-12 08:02:00	10.386033
2018-08-12 08:03:00	9.068647
2018-08-12 08:04:00	11.204582
2018-08-12 08:05:00	9.076552
2018-08-12 08:06:00	9.767191
2018-08-12 08:07:00	9.117590
2018-08-12 08:08:00	9.202167
2018-08-12 08:09:00	10.829297
2018-08-12 08:10:00	9.940771

In [41]:
```
df2 = df['2018-08-12']
df2.head()
```

Out[41]:

	cpu
time	
2018-08-12 00:00:00	9.773730
2018-08-12 00:01:00	8.769905
2018-08-12 00:02:00	8.974288
2018-08-12 00:03:00	10.628897
2018-08-12 00:04:00	9.055220

In [42]:
```
df2.tail()
```

Out[42]:

	cpu
time	
2018-08-12 23:55:00	11.823216
2018-08-12 23:56:00	11.391370
2018-08-12 23:57:00	9.713983
2018-08-12 23:58:00	10.756149
2018-08-12 23:59:00	9.155902

In [43]:
```
df2 = df.groupby(df.index.date).mean()
print(df2.head())
print(df2.tail())
```

```
                 cpu
2018-08-12  10.026842
2018-08-13  10.021321
2018-08-14   9.980860
2018-08-15  10.011482
2018-08-16   9.998402
                 cpu
2018-12-24   9.972200
2018-12-25  10.033823
2018-12-26  10.022897
2018-12-27   9.992240
2018-12-28   9.977914
```

In [44]:
```
df2 = df.groupby(df.index.hour).mean()
print(df2.head())
print(df2.tail())
```

```
            cpu
time
0     10.001204
1      9.998397
2      9.997178
3     10.001992
4     10.006139
            cpu
time
19     9.992884
20     9.982940
21    10.016058
22     9.989161
23    10.004904
```

In [45]:
```
df2 = df.groupby(df.index.week).mean()
print(df2.head())
print(df2.tail())
```

```
            cpu
time
32    10.026842
33    10.001378
34     9.996593
35    10.008314
36     9.989986
            cpu
time
48    10.006039
49     9.990210
50     9.993180
51     9.980866
52    10.000312
```

resample

In [46]:
```
df2 = df.resample('90S').mean()              # 重新取样，时间间隔为90秒了
```

In [47]:
```
df2.head()
```

Out[47]:

	cpu
time	
2018-08-12 00:00:00	9.271818
2018-08-12 00:01:30	8.974288
2018-08-12 00:03:00	9.842058
2018-08-12 00:04:30	9.176499
2018-08-12 00:06:00	10.425718

In [48]:
```
df2 = df.resample('5T').max()
df2.head()
```

Out[48]:

	cpu
time	
2018-08-12 00:00:00	10.628897
2018-08-12 00:05:00	10.740297
2018-08-12 00:10:00	10.542998
2018-08-12 00:15:00	10.864424
2018-08-12 00:20:00	10.435594