Tashu-analysis

2017년 3월 6일

1 타슈 데이터 분석

```
In [1]: %matplotlib inline
        import pandas as pd
        import pylab
        import datetime
        import numpy
        import matplotlib.image
        import matplotlib.pyplot as plt
```

1.1 Open data with csv file

```
In [2]: #rent = pd.read_csv('./2013_01.csv', parse_dates=[2,4])
        rent = pd.read_csv('2013_01.csv', parse_dates=[2,4])
```

```
In [3]: rent[:10]
```

```
Out[3]:   member  rent_station           rent_date  return_station         return_date
        0      No          43.0 2013-01-01 05:56:03            34.0 2013-01-01 06:02:17
        1      No          97.0 2013-01-01 06:04:00             NaN 2013-01-01 10:20:37
        2      No           2.0 2013-01-01 06:04:06            10.0 2013-01-01 06:18:59
        3      No         106.0 2013-01-01 10:53:05           105.0 2013-01-01 10:57:43
        4     Yes           4.0 2013-01-01 11:22:23             4.0 2013-01-01 12:17:53
        5      No          21.0 2013-01-01 11:39:53           105.0 2013-01-01 11:49:43
        6      No          90.0 2013-01-01 12:08:33            91.0 2013-01-01 12:51:36
        7      No          13.0 2013-01-01 13:14:29            30.0 2013-01-01 13:30:39
        8     Yes           1.0 2013-01-01 13:37:42             1.0 2013-01-01 13:38:15
        9     Yes           1.0 2013-01-01 13:38:13             2.0 2013-01-01 15:09:58
```

```
In [4]: rent['sub'] = rent.return_date - rent.rent_date
        rent['sub'].astype('timedelta64[s]')
        rent[:10]
```

```
Out[4]:   member  rent_station           rent_date  return_station  \
        0     No          43.0 2013-01-01 05:56:03            34.0
        1     No          97.0 2013-01-01 06:04:00             NaN
        2     No           2.0 2013-01-01 06:04:06            10.0
        3     No         106.0 2013-01-01 10:53:05           105.0
        4    Yes           4.0 2013-01-01 11:22:23             4.0
        5     No          21.0 2013-01-01 11:39:53           105.0
        6     No          90.0 2013-01-01 12:08:33            91.0
        7     No          13.0 2013-01-01 13:14:29            30.0
        8    Yes           1.0 2013-01-01 13:37:42             1.0
        9    Yes           1.0 2013-01-01 13:38:13             2.0

                  return_date      sub
        0 2013-01-01 06:02:17 00:06:14
        1 2013-01-01 10:20:37 04:16:37
        2 2013-01-01 06:18:59 00:14:53
        3 2013-01-01 10:57:43 00:04:38
        4 2013-01-01 12:17:53 00:55:30
        5 2013-01-01 11:49:43 00:09:50
        6 2013-01-01 12:51:36 00:43:03
        7 2013-01-01 13:30:39 00:16:10
        8 2013-01-01 13:38:15 00:00:33
        9 2013-01-01 15:09:58 01:31:45
```

```
In [5]: ts = pd.Series(1, rent['rent_date'])
        ts[:10]
```

```
Out[5]: rent_date
        2013-01-01 05:56:03    1
        2013-01-01 06:04:00    1
        2013-01-01 06:04:06    1
        2013-01-01 10:53:05    1
        2013-01-01 11:22:23    1
        2013-01-01 11:39:53    1
        2013-01-01 12:08:33    1
        2013-01-01 13:14:29    1
        2013-01-01 13:37:42    1
        2013-01-01 13:38:13    1
        dtype: int64
```

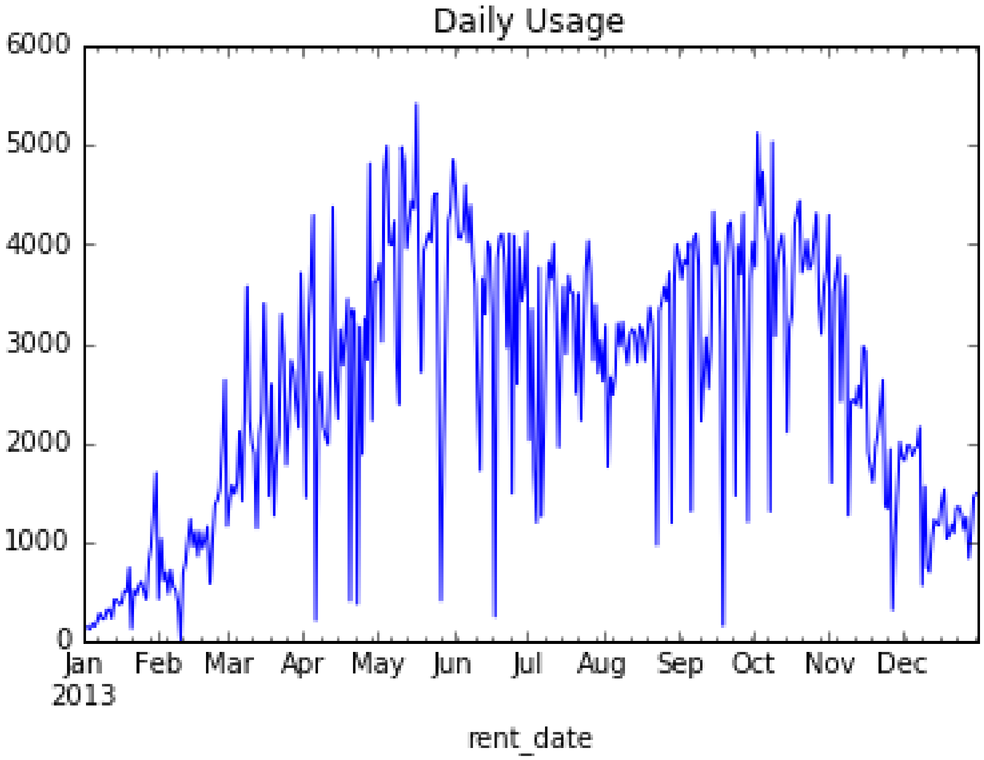
```
In [6]: #daily_ts = ts.resample('D', how='count')
        daily_ts = ts.resample('D').count()

        daily_ts[:10]
```

```
Out[6]: rent_date
        2013-01-01     70
        2013-01-02    153
        2013-01-03    146
        2013-01-04    129
        2013-01-05    183
        2013-01-06    155
        2013-01-07    209
        2013-01-08    292
        2013-01-09    234
        2013-01-10    233
        Freq: D, dtype: int64
```

```
In [7]: #monthly_ts = ts.resample('M', how='count')
        monthly_ts = ts.resample('M').count()
        monthly_ts[:10]
```

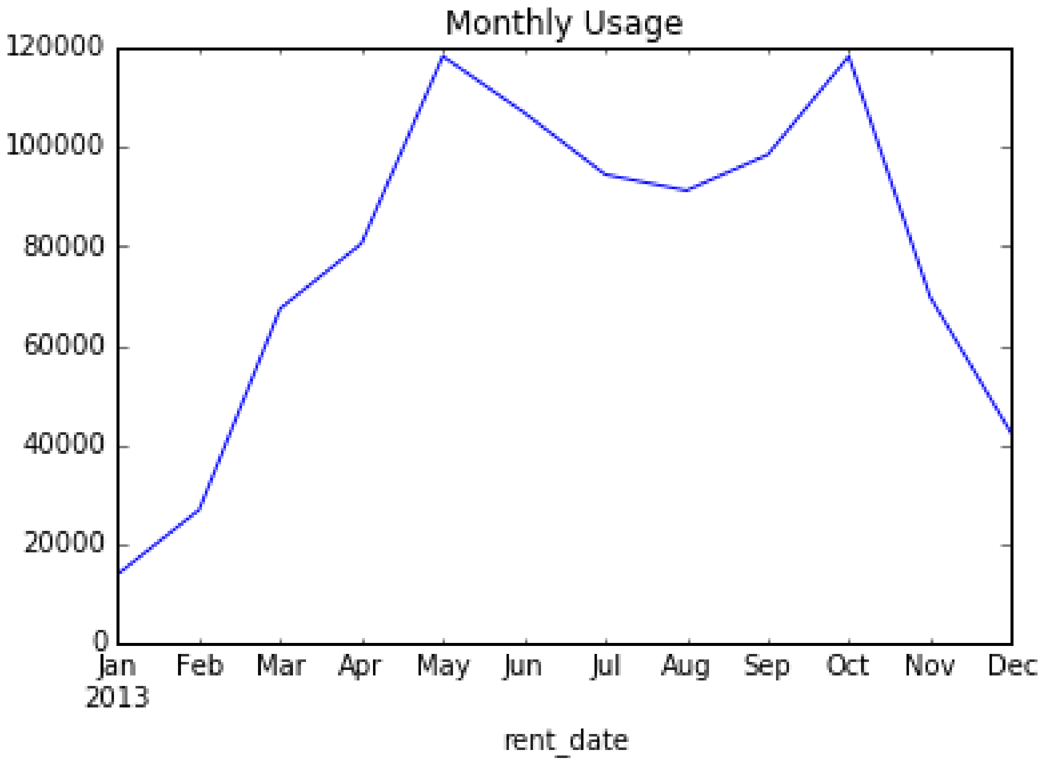
```
Out[7]: rent_date
        2013-01-31     14084
        2013-02-28     27028
        2013-03-31     67550
        2013-04-30     80637
        2013-05-31    118336
        2013-06-30    107004
        2013-07-31     94534
        2013-08-31     91314
        2013-09-30     98520
        2013-10-31    118318
        Freq: M, dtype: int64
```

```
In [8]: plt.title('Daily Usage')
        daily_ts.plot()
        plt.show()
```



```
In [9]: pylab.savefig('daily.pdf')

<matplotlib.figure.Figure at 0x11137f050>
```

```
In [10]: plt.title('Monthly Usage')
         monthly_ts.plot()

Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x10f11b190>
```



```
In [11]: pylab.savefig('monthly.pdf')
```

```
<matplotlib.figure.Figure at 0x118ec1b10>
```

1.2 weekly stat

```
In [12]: rent['weekday'] = pd.DatetimeIndex(rent['rent_date']).weekday
         rent[:10]
```

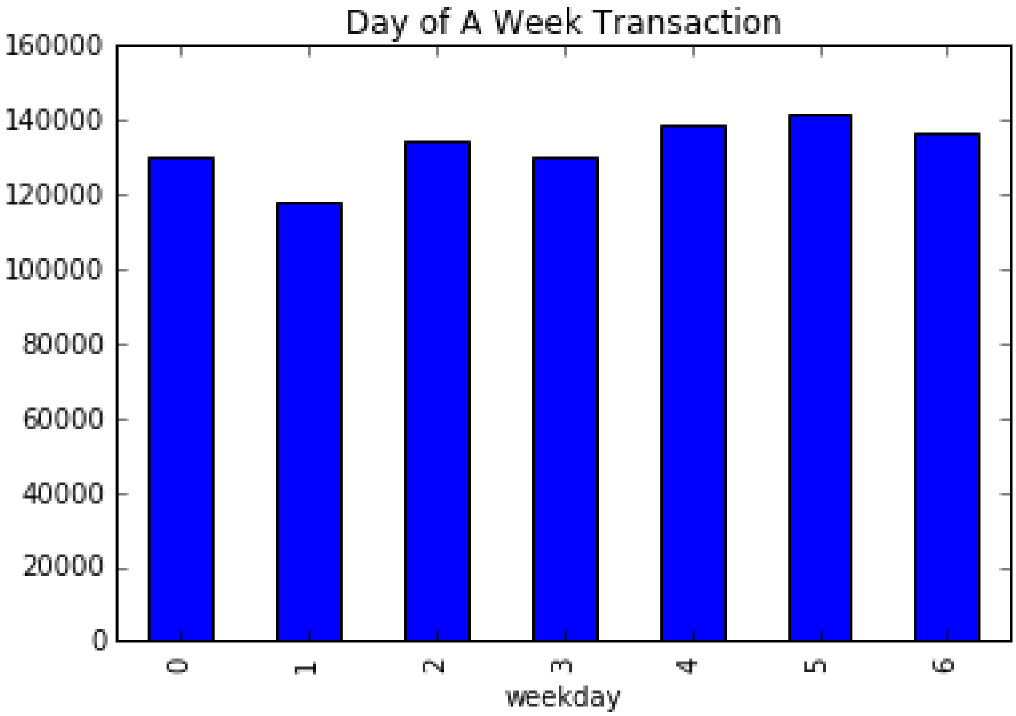
```
Out[12]:   member  rent_station           rent_date  return_station  \
         0     No          43.0 2013-01-01 05:56:03            34.0
         1     No          97.0 2013-01-01 06:04:00             NaN
         2     No           2.0 2013-01-01 06:04:06            10.0
         3     No         106.0 2013-01-01 10:53:05           105.0
         4    Yes           4.0 2013-01-01 11:22:23             4.0
         5     No          21.0 2013-01-01 11:39:53           105.0
         6     No          90.0 2013-01-01 12:08:33            91.0
```

```
7     No         13.0 2013-01-01 13:14:29            30.0
8    Yes          1.0 2013-01-01 13:37:42             1.0
9    Yes          1.0 2013-01-01 13:38:13             2.0

          return_date      sub  weekday
0 2013-01-01 06:02:17 00:06:14        1
1 2013-01-01 10:20:37 04:16:37        1
2 2013-01-01 06:18:59 00:14:53        1
3 2013-01-01 10:57:43 00:04:38        1
4 2013-01-01 12:17:53 00:55:30        1
5 2013-01-01 11:49:43 00:09:50        1
6 2013-01-01 12:51:36 00:43:03        1
7 2013-01-01 13:30:39 00:16:10        1
8 2013-01-01 13:38:15 00:00:33        1
9 2013-01-01 15:09:58 01:31:45        1
```

```
In [13]: rent_weekday = rent.groupby('weekday').rent_station.count()
         rent_weekday
```

```
Out[13]: weekday
         0    130406
         1    117726
         2    134451
         3    130035
         4    138440
         5    141733
         6    136470
         Name: rent_station, dtype: int64
```

```
In [14]: plt.xticks((0,1,2,3,4,5,6),('Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'S
         plt.title('Day of A Week Transaction')
         #rent_weekday.plot()
         rent_weekday.plot(kind='bar'); plt.axhline(0, color='k')
         #plt.hist(rent_weekday)
         plt.show()
```



```
In [15]: from numpy import *

In [16]: rent[:10]

Out[16]:   member  rent_station           rent_date  return_station  \
         0      No          43.0 2013-01-01 05:56:03            34.0
         1      No          97.0 2013-01-01 06:04:00             NaN
         2      No           2.0 2013-01-01 06:04:06            10.0
         3      No         106.0 2013-01-01 10:53:05           105.0
         4     Yes           4.0 2013-01-01 11:22:23             4.0
         5      No          21.0 2013-01-01 11:39:53           105.0
         6      No          90.0 2013-01-01 12:08:33            91.0
         7      No          13.0 2013-01-01 13:14:29            30.0
         8     Yes           1.0 2013-01-01 13:37:42             1.0
         9     Yes           1.0 2013-01-01 13:38:13             2.0

                   return_date      sub  weekday
         0 2013-01-01 06:02:17 00:06:14        1
         1 2013-01-01 10:20:37 04:16:37        1
         2 2013-01-01 06:18:59 00:14:53        1
         3 2013-01-01 10:57:43 00:04:38        1
         4 2013-01-01 12:17:53 00:55:30        1
         5 2013-01-01 11:49:43 00:09:50        1
         6 2013-01-01 12:51:36 00:43:03        1
         7 2013-01-01 13:30:39 00:16:10        1
         8 2013-01-01 13:38:15 00:00:33        1
         9 2013-01-01 15:09:58 01:31:45        1
```

1.3 integer conversion

```
In [17]: import warnings
         warnings.filterwarnings('ignore')

         whereAreNaNs = isnan(rent.return_station)
         rent.return_station[whereAreNaNs] = 0
         whereAreNaNs = isnan(rent.rent_station)
         rent.rent_station[whereAreNaNs] = 0



         rent['rent_station'] = rent['rent_station'].astype('int')
         rent['return_station'] = rent['return_station'].astype('int')
         rent[:10]
```

```
Out[17]:   member  rent_station           rent_date  return_station  \
         0     No            43 2013-01-01 05:56:03              34
         1     No            97 2013-01-01 06:04:00               0
         2     No             2 2013-01-01 06:04:06              10
         3     No           106 2013-01-01 10:53:05             105
         4    Yes             4 2013-01-01 11:22:23               4
         5     No            21 2013-01-01 11:39:53             105
         6     No            90 2013-01-01 12:08:33              91
         7     No            13 2013-01-01 13:14:29              30
         8    Yes             1 2013-01-01 13:37:42               1
         9    Yes             1 2013-01-01 13:38:13               2
```

```
          return_date      sub  weekday
0 2013-01-01 06:02:17 00:06:14        1
1 2013-01-01 10:20:37 04:16:37        1
2 2013-01-01 06:18:59 00:14:53        1
3 2013-01-01 10:57:43 00:04:38        1
4 2013-01-01 12:17:53 00:55:30        1
5 2013-01-01 11:49:43 00:09:50        1
6 2013-01-01 12:51:36 00:43:03        1
7 2013-01-01 13:30:39 00:16:10        1
8 2013-01-01 13:38:15 00:00:33        1
9 2013-01-01 15:09:58 01:31:45        1
```

1.4 Top 10 stations ?

```
In [18]: rent_station_cnt = rent.groupby('rent_station').rent_station.count()
         rent_station_cnt.sort_values(inplace=True, ascending=False)

         rent_station_cnt[:10]
```

```
Out[18]: rent_station
         3      49416
         56     30383
         17     26262
         31     25968
         33     22792
         32     22196
         14     18611
         60     17243
         55     17060
         8      16820
         Name: rent_station, dtype: int64
```

```
In [ ]:
```