数据挖掘第二次作业 q1 报告

1651718 方沛

题目要求

数据预处理：首先要求对每个商品编号/商品类别结构/品牌编号等按照购买时间对购买数量进行 汇总求和，包括按天、周、约为周期分别形成该商品编号/商品类别结构/品牌编号对应购买数量 汇总的每天。

代码设计

预处理数据

读入数据，生成四个品类的新特征

```
dataset = pd.read_csv("trade_new.csv").fillna(0)
dataset['sldatime']=pd.to_datetime(dataset['sldatime'])
dataset['sldatime']=[datetime.strftime(x,'%Y-%m-%d') for x in dataset['sldatime']]
# 转化格式并截取
dataset['pluno']=dataset['pluno'].astype("str")
dataset['pl1']=[x[:2] for x in dataset['pluno']]
dataset['pl2']=[x[:3] for x in dataset['pluno']]
dataset['pl3']=[x[:4] for x in dataset['pluno']]
dataset['pl4']=[x[:5] for x in dataset['pluno']]
```

主要函数

```
'''
@descrption: 生成时序数据一个属性对另一个属性的时序数据
@params:
    - period: 按照什么时间段进行统计. 'M',月; 'W',星期; 'D',日;
    - dataset: 目标数据集
    - metric: 被统计的单位,这里是 'pluno','bndno' 或者 四个品类结构.
    - target: 统计指标, 这里是'qty'
@output: 返回的经处理数据集
'''
def generate_time_series(period,dataset,metric,target):
    pluno_series = dict()
    dataset_time=dataset.to_period(period)
    dataset_time.sort_index(inplace=True)
    for freq in np.unique(dataset_time.index):
        orders=dataset_time.loc[freq,[metric,target]]
        pluno_series[freq]=dict()
        for index,row in orders.iterrows():
            pluno_series[freq][row[metric]]=row[target] if row[target] not in
pluno_series[freq] else pluno_series[freq][row[metric]]+row[target]
    return pd.DataFrame(pluno_series).fillna(0)
```

运行结果

pluno 按月统计

In [7]:
```
#按月
generate_time_series("M",dataset,'pluno','qty').head(10)
```

Out[7]:

	2016-02	2016-03	2016-04	2016-05	2016-06	2016-07
10000000	1.0	1.0	0.0	0.0	1.0	0.0
10000003	0.0	1.0	0.0	0.0	0.0	1.0
10000004	1.0	0.0	1.0	1.0	1.0	1.0
10000005	1.0	0.0	0.0	1.0	1.0	1.0
10000006	1.0	1.0	1.0	1.0	1.0	1.0
10000007	1.0	1.0	1.0	0.0	1.0	0.0
10000009	1.0	1.0	0.0	1.0	1.0	0.0
10000010	1.0	0.0	0.0	1.0	1.0	1.0
10000011	1.0	1.0	1.0	1.0	1.0	1.0
10000012	1.0	1.0	1.0	1.0	0.0	1.0

pluno按周统计

```
#按星期
generate_time_series("W",dataset,'pluno','qty').head(10)
```

	2016-02-01/2016-02-07	2016-02-08/2016-02-14	2016-02-15/2016-02-21	2016-02-22/2016-02-28	2016-02-29/2016-03-06	2016-03-07/2016-03-13	2016-03-14/2016-03-20	2016-03-21/2016-03-27	2016-03-28/2016-04-03	2016-04-04/2016-04-10	...	2016-05-23/2016-05-29	2016-05-30/2016-06-05	2016-06-06/2016-06-12	2016-06-13/2016-06-19	2016-06-20/2016-06-2
10000000	0.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	1.0	0.
10000003	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.
10000004	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	...	0.0	0.0	0.0	1.0	0.
10000005	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	1.0	0.0	0.0	1.0	0.
10000006	0.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0	1.0	...	1.0	1.0	0.0	1.0	1.
10000007	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	1.
10000009	1.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	...	1.0	1.0	0.0	0.0	0.
10000010	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	1.
10000011	1.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	...	1.0	0.0	1.0	1.0	0.
10000012	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.

10 rows × 26 columns

pluno按日统计

In [9]:
```
#按日
generate_time_series("D",dataset,'pluno','qty').head(10)
```

Out[9]:

	2016-02-01	2016-02-02	2016-02-03	2016-02-04	2016-02-05	2016-02-06	2016-02-07	2016-02-08	2016-02-09	2016-02-10	...	2016-07-22	2016-07-23	2016-07-24	2016-07-25	2016-07-26	2016-07-27	2016-07-28	2016-07-29	2016-07-30	2016-07-31
10000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10000003	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10000004	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10000005	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10000006	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10000007	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10000009	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10000010	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0
10000011	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10000012	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

10 rows × 182 columns

一 级 品 类 按 月 统 计

In [10]:
```
#按月
generate_time_series("M",dataset,'pl1','qty').head(10)
```

Out[10]:

	2016-02	2016-03	2016-04	2016-05	2016-06	2016-07
10	10.000	1.000	1.000	1.000	1.000	1.000
11	1.000	1.000	1.000	1.000	1.000	1.000
14	0.434	1.000	1.000	1.000	1.000	1.000
15	1.000	1.000	1.000	1.000	2.000	1.000
20	0.514	242.000	0.010	0.003	0.124	0.800
21	0.426	0.124	0.900	1.054	0.764	0.244
22	0.946	0.396	0.666	392.000	0.760	0.634
23	1.000	1.000	1.000	10.000	10.000	1.000
24	0.466	0.304	0.378	1.534	0.100	0.990
25	0.304	0.200	2.000	0.112	0.298	0.182

一级品类按周统计

```
#按周
generate_time_series("W",dataset,'pl1','qty').head(10)
```

	2016-02-01/2016-02-07	2016-02-08/2016-02-14	2016-02-15/2016-02-21	2016-02-22/2016-02-28	2016-02-29/2016-03-06	2016-03-07/2016-03-13	2016-03-14/2016-03-20	2016-03-21/2016-03-27	2016-03-28/2016-04-03	2016-04-04/2016-04-10	...	2016-05-23/2016-05-29	2016-05-30/2016-06-05	2016-06-06/2016-06-12	2016-06-13/2016-06-19	2016-06-20/2016-06-26	2 27/2 0
10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	...	1.000	1.000	1.000	1.000	1.000	1
11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	...	1.000	1.000	1.000	1.000	1.000	1
14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	...	1.000	1.000	1.000	1.000	1.000	1
15	1.000	1.000	1.000	1.000	2.000	1.000	1.000	6.000	6.000	1.000	...	6.000	1.000	1.000	1.000	6.000	1
20	1.766	378.000	0.514	0.394	0.250	0.162	0.004	0.374	0.538	0.260	...	0.124	0.948	0.118	0.344	0.948	0
21	3.100	0.510	2.216	0.170	0.284	0.790	0.001	0.774	1.228	0.092	...	0.756	1.062	0.616	0.278	0.846	0
22	0.448	0.298	0.002	0.001	0.516	0.794	1.042	0.196	5.000	0.338	...	1.348	1.666	0.972	2.298	1.092	0
23	0.606	0.001	1.000	1.000	1.000	0.001	0.254	1.000	1.000	6.000	...	10.000	0.982	10.000	1.000	0.306	0
24	0.932	0.434	0.358	0.166	0.350	0.518	0.534	0.298	0.258	0.286	...	0.218	0.278	0.642	0.594	1.340	0
25	0.352	2.000	3.000	1.508	0.244	0.004	0.200	4.000	0.312	0.200	...	0.128	2.000	0.004	0.300	2.000	2

10 rows × 26 columns

一级品类按日统计

```
#按日
generate_time_series("D",dataset,'pl1','qty').head(10)
```

	2016-02-01	2016-02-02	2016-02-03	2016-02-04	2016-02-05	2016-02-06	2016-02-07	2016-02-08	2016-02-09	2016-02-10	...	2016-07-22	2016-07-23	2016-07-24	2016-07-25	2016-07-26	2016-07-27	2016-07-28	2016-07-29	2016-07-30	2016-07-31
10	1.000	1.000	6.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	...	1.000	1.000	1.000	1.000	1.000	6.000	1.000	1.000	1.000	1.000
11	1.000	1.000	1.000	1.000	1.000	3.000	1.000	1.000	0.000	1.000	...	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
14	1.000	1.000	1.000	1.000	0.434	1.000	1.000	1.000	1.000	1.000	...	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.356	1.000
15	0.310	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	6.000	...	1.000	12.000	1.000	2.000	1.000	1.000	1.000	1.000	6.000	1.000
20	1.518	0.000	0.554	0.704	0.552	0.786	0.000	0.000	0.000	434.000	...	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.045
21	0.448	0.426	0.102	0.001	0.000	0.194	0.000	0.000	0.000	0.510	...	0.122	0.298	0.428	0.000	0.000	0.270	0.000	0.000	0.694	0.606
22	0.430	1.010	0.642	1.066	0.732	0.598	1.512	0.404	0.244	0.280	...	0.552	1.282	0.900	1.302	1.088	0.001	0.738	0.432	0.966	0.900
23	1.000	0.208	1.000	10.000	0.001	10.000	1.000	0.430	10.000	0.448	...	0.432	0.001	15.000	10.000	1.000	1.000	0.001	0.284	0.001	0.156
24	0.438	1.098	0.240	0.390	0.836	0.190	0.482	0.452	0.840	0.606	...	0.384	0.576	0.600	3.156	1.114	0.412	0.000	0.990	0.448	0.396
25	0.352	0.290	2.000	0.200	0.308	2.000	0.020	0.000	0.001	0.004	...	3.000	4.000	0.546	0.000	0.001	0.200	0.200	3.000	2.000	0.005

10 rows × 182 columns

二级品类按月统计

In [11]:
```
generate_time_series("M",dataset,'p12','qty').head(10)
```

Out[11]:

	2016-02	2016-03	2016-04	2016-05	2016-06	2016-07
100	1.0	1.0	1.0	1.0	1.0	1.0
101	1.0	1.0	1.0	1.0	1.0	1.0
102	1.0	1.0	1.0	1.0	1.0	1.0
103	2.0	1.0	1.0	1.0	1.0	1.0
104	1.0	1.0	1.0	2.0	1.0	1.0
105	10.0	1.0	1.0	10.0	1.0	0.0
106	1.0	0.0	0.0	0.0	0.0	0.0
107	0.0	0.0	0.0	18.0	1.0	0.0
110	1.0	1.0	1.0	1.0	1.0	1.0
111	1.0	1.0	1.0	1.0	1.0	1.0

二级品类按周统计

In [15]:
```
generate_time_series("W",dataset,'p12','qty').head(10)
```

Out[15]:

	2016-02-01/2016-02-07	2016-02-08/2016-02-14	2016-02-15/2016-02-21	2016-02-22/2016-02-28	2016-02-29/2016-03-06	2016-03-07/2016-03-13	2016-03-14/2016-03-20	2016-03-21/2016-03-27	2016-03-28/2016-04-03	2016-04-04/2016-04-10	...	2016-05-23/2016-05-29	2016-05-30/2016-06-05	2016-06-06/2016-06-12	2016-06-13/2016-06-19	2016-06-20/2016-06-26	27
100	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	...	1.0	1.0	1.0	1.0	1.0	
101	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	...	1.0	1.0	1.0	1.0	1.0	
102	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	6.0	...	1.0	1.0	1.0	0.0	1.0	
103	0.0	24.0	2.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	...	24.0	1.0	1.0	1.0	1.0	
104	1.0	1.0	1.0	3.0	1.0	0.0	1.0	1.0	1.0	0.0	...	1.0	1.0	1.0	1.0	1.0	
105	10.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	10.0	1.0	...	1.0	0.0	0.0	1.0	0.0	
106	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	
107	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	12.0	1.0	
110	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	...	1.0	1.0	1.0	1.0	1.0	
111	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	39.0	1.0	...	1.0	148.0	1.0	1.0	1.0	

10 rows × 26 columns

二级品类按日统计

```
generate_time_series("D",dataset,'p12','qty').head(10)
```

:

	2016-02-01	2016-02-02	2016-02-03	2016-02-04	2016-02-05	2016-02-06	2016-02-07	2016-02-08	2016-02-09	2016-02-10	...	2016-07-22	2016-07-23	2016-07-24	2016-07-25	2016-07-26	2016-07-27	2016-07-28	2016-07-29	2016-07-30	201 07-
100	1.0	1.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	...	1.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	1.0	1
101	6.0	1.0	6.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	...	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1
102	1.0	1.0	4.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	...	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1
103	0.0	0.0	0.0	0.0	0.0	0.0	0.0	24.0	0.0	0.0	...	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0
104	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	...	1.0	0.0	1.0	0.0	1.0	6.0	1.0	1.0	0.0	0
105	0.0	0.0	0.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
106	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
107	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
110	1.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.0	...	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	1
111	0.0	1.0	0.0	0.0	0.0	1.0	1.0	1.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1

10 rows × 182 columns

三级品类按月统计

In [12]:
```
generate_time_series("M",dataset,'p13','qty').head(10)
```

Out[12]:

	2016-02	2016-03	2016-04	2016-05	2016-06	2016-07
1000	1.0	1.0	1.0	1.0	1.0	1.0
1010	1.0	0.0	1.0	1.0	1.0	1.0
1011	1.0	1.0	1.0	1.0	1.0	1.0
1012	0.0	0.0	0.0	0.0	1.0	0.0
1013	1.0	1.0	1.0	1.0	1.0	1.0
1014	1.0	1.0	1.0	1.0	1.0	1.0
1015	1.0	1.0	1.0	1.0	1.0	1.0
1020	1.0	1.0	1.0	1.0	1.0	1.0
1030	2.0	1.0	1.0	1.0	12.0	1.0
1031	0.0	1.0	1.0	1.0	1.0	1.0

三级品类按周统计

```
generate_time_series("W",dataset,'p13','qty').head(10)
```

	2016-02-01/2016-02-07	2016-02-08/2016-02-14	2016-02-15/2016-02-21	2016-02-22/2016-02-28	2016-02-29/2016-03-06	2016-03-07/2016-03-13	2016-03-14/2016-03-20	2016-03-21/2016-03-27	2016-03-28/2016-04-03	2016-04-04/2016-04-10	...	2016-05-23/2016-05-29	2016-05-30/2016-06-05	2016-06-06/2016-06-12	2016-06-13/2016-06-19	2016-06-20/2016-06-26	2
1000	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	...	1.0	1.0	1.0	1.0	1.0	
1010	1.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	...	0.0	1.0	1.0	1.0	1.0	
1011	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	...	1.0	1.0	1.0	1.0	1.0	
1012	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	
1013	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	...	1.0	1.0	1.0	1.0	1.0	
1014	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	...	1.0	1.0	1.0	1.0	1.0	
1015	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	...	1.0	1.0	1.0	1.0	1.0	
1020	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	6.0	...	1.0	1.0	1.0	0.0	1.0	
1030	0.0	24.0	2.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	...	24.0	1.0	0.0	0.0	0.0	
1031	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.0	...	0.0	0.0	1.0	1.0	1.0	

10 rows × 26 columns

三级品类按日统计

```
generate_time_series("D",dataset,'p13','qty').head(10)
```

	2016-02-01	2016-02-02	2016-02-03	2016-02-04	2016-02-05	2016-02-06	2016-02-07	2016-02-08	2016-02-09	2016-02-10	...	2016-07-22	2016-07-23	2016-07-24	2016-07-25	2016-07-26	2016-07-27	2016-07-28	2016-07-29	2016-07-30	2016-07-31
1000	1.0	1.0	1.0	0.0	0.0	1.0	1.0	1.0	1.0	0.0	...	1.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0
1010	1.0	0.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	...	0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0
1011	6.0	1.0	6.0	0.0	10.0	1.0	1.0	1.0	1.0	1.0	...	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0
1012	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1013	1.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	...	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0
1014	0.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0	1.0	...	1.0	1.0	0.0	1.0	1.0	1.0	1.0	2.0	0.0	1.0
1015	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0	1.0	...	1.0	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0
1020	1.0	1.0	4.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	...	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
1030	0.0	0.0	0.0	0.0	0.0	0.0	0.0	24.0	0.0	0.0	...	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1031	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

四级品类按月统计

```
In [13]:
1 generate_time_series("M",dataset,'p14','qty').head(10)
```

Out[13]:

	2016-02	2016-03	2016-04	2016-05	2016-06	2016-07
10000	1.0	1.0	1.0	1.0	1.0	1.0
10001	1.0	1.0	1.0	1.0	1.0	1.0
10002	1.0	1.0	1.0	1.0	1.0	1.0
10008	1.0	0.0	1.0	1.0	1.0	0.0
10100	1.0	0.0	0.0	1.0	1.0	1.0
10101	1.0	0.0	0.0	0.0	0.0	0.0
10102	1.0	0.0	0.0	0.0	1.0	0.0
10103	1.0	0.0	0.0	1.0	0.0	1.0
10109	1.0	0.0	1.0	1.0	1.0	1.0
10110	1.0	1.0	1.0	1.0	1.0	1.0

四级品类按周统计

```
In [17]:
1 generate_time_series("W",dataset,'p14','qty').head(10)
```

Out[17]:

	2016-02-01/2016-02-07	2016-02-08/2016-02-14	2016-02-15/2016-02-21	2016-02-22/2016-02-28	2016-02-29/2016-03-06	2016-03-07/2016-03-13	2016-03-14/2016-03-20	2016-03-21/2016-03-27	2016-03-28/2016-04-03	2016-04-04/2016-04-10	...	2016-05-23/2016-05-29	2016-05-30/2016-06-05	2016-06-06/2016-06-12	2016-06-13/2016-06-19	2016-06-20/2016-06-26
10000	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	...	1.0	1.0	1.0	1.0	1.0
10001	1.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0	...	0.0	1.0	1.0	1.0	0.0
10002	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	...	1.0	1.0	1.0	1.0	1.0
10008	0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	...	1.0	1.0	1.0	1.0	0.0
10100	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0	1.0	1.0
10101	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
10102	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	1.0
10103	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
10109	1.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	...	0.0	1.0	0.0	1.0	1.0
10110	1.0	2.0	1.0	0.0	0.0	0.0	1.0	1.0	1.0	1.0	...	1.0	1.0	0.0	1.0	1.0

10 rows × 26 columns

四级品类按日统计

```
generate_time_series("D",dataset,'p14','qty').head(10)
```

	2016-02-01	2016-02-02	2016-02-03	2016-02-04	2016-02-05	2016-02-06	2016-02-07	2016-02-08	2016-02-09	2016-02-10	...	2016-07-22	2016-07-23	2016-07-24	2016-07-25	2016-07-26	2016-07-27	2016-07-28	2016-07-29	2016-07-30	2016-07-31
10000	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	...	1.0	1.0	1.0	0.0	0.0	0.0	1.0	1.0	1.0	0.0
10001	1.0	1.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10002	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	...	1.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	1.0	1.0
10008	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
10101	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10102	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10103	1.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10109	1.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0
10110	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	...	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0

10 rows × 182 columns