q1报告

作者：王星洲

学号：1652977

步骤

1. 读取数据，并将日期规范化

```
df.loc[:,"sldatime"] = df.loc[:,"sldatime"].apply(lambda x:x[:10]).tolist()
df
```

2. 将每个日期对应的周和月求出来

```
def judgeDate(a):
    delta = (datetime.datetime(int(a[0:4]),int(a[5:7]),int(a[8:10])) - datetime.datetime(2016,2,1)).days
    week = int(delta / 7)
    month = int(a[5:7])
    return (week, month)
```

3. 舍弃不需要的数据

```
a = df1.loc[df1["kind1"]== 22]
b = df1.loc[df1["kind1"]== 23]
c = df1.loc[df1["kind1"]== 25]
d = df1.loc[df1["kind1"]== 27]
e = df1.loc[df1["kind4"]== 15000]
df1 = a.append(b).append(c).append(d).append(e)
df1 = df1.reset_index(drop=True)
df1
```

4. 生成所需的时间序列，这里以商品-日期为例

```
df11 = pd.DataFrame([],columns=plu_list.to_list(),index=sldatime_list)
for i in range(df1.shape[0]):
    df11.loc[df1.loc[i,"sldatime"],df1.loc[i,"pluno"]]=df1.loc[i,"qty"]
df11 = df11.fillna(0)
df11
```

输出

日期/周/月 * 商品号/品牌号/四级品类的时间序列

共18个csv文件