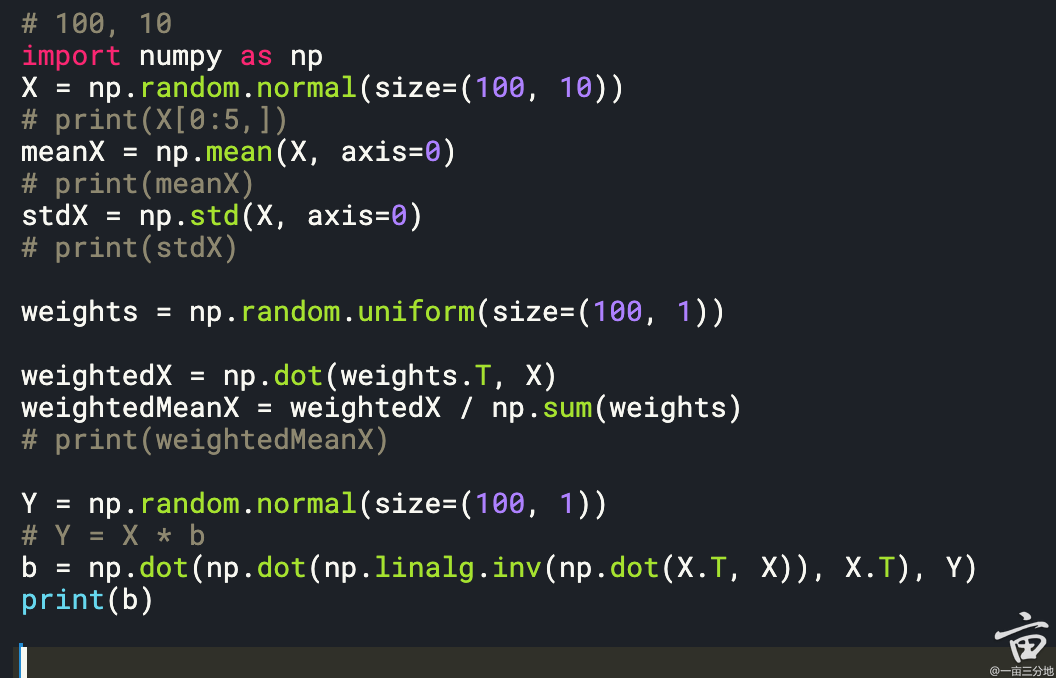
**np.random相关操作**

groupby object .to\_frame()

to\_datetime()

一些基本numpy matrix 操作

<https://www.1point3acres.com/bbs/thread-680805-1-1.html>



Q5: weighted monthly average market cap per company

The way C calculates and populates the monthly total volume (to be used the denominator for weights) for each company is pretty clean

```

df\_company['total\_dollar\_volume'] = df\_company.groupby(['year', 'month','company\_id'])['dollar\_volume'].transform('sum')

df\_company['weight'] =df\_company['dollar\_volume'] / df\_company['total\_dollar\_volume']

```

Q5: subset the top 5 company

This is also very creative.

```

# top 5 by month

df\_sorted = df\_weighted.sort\_values(by=['year', 'month','weighted\_mkt\_cap'], ascending=[True, True, False])

df\_top = df\_sorted.groupby(['year', 'month']).head(5)

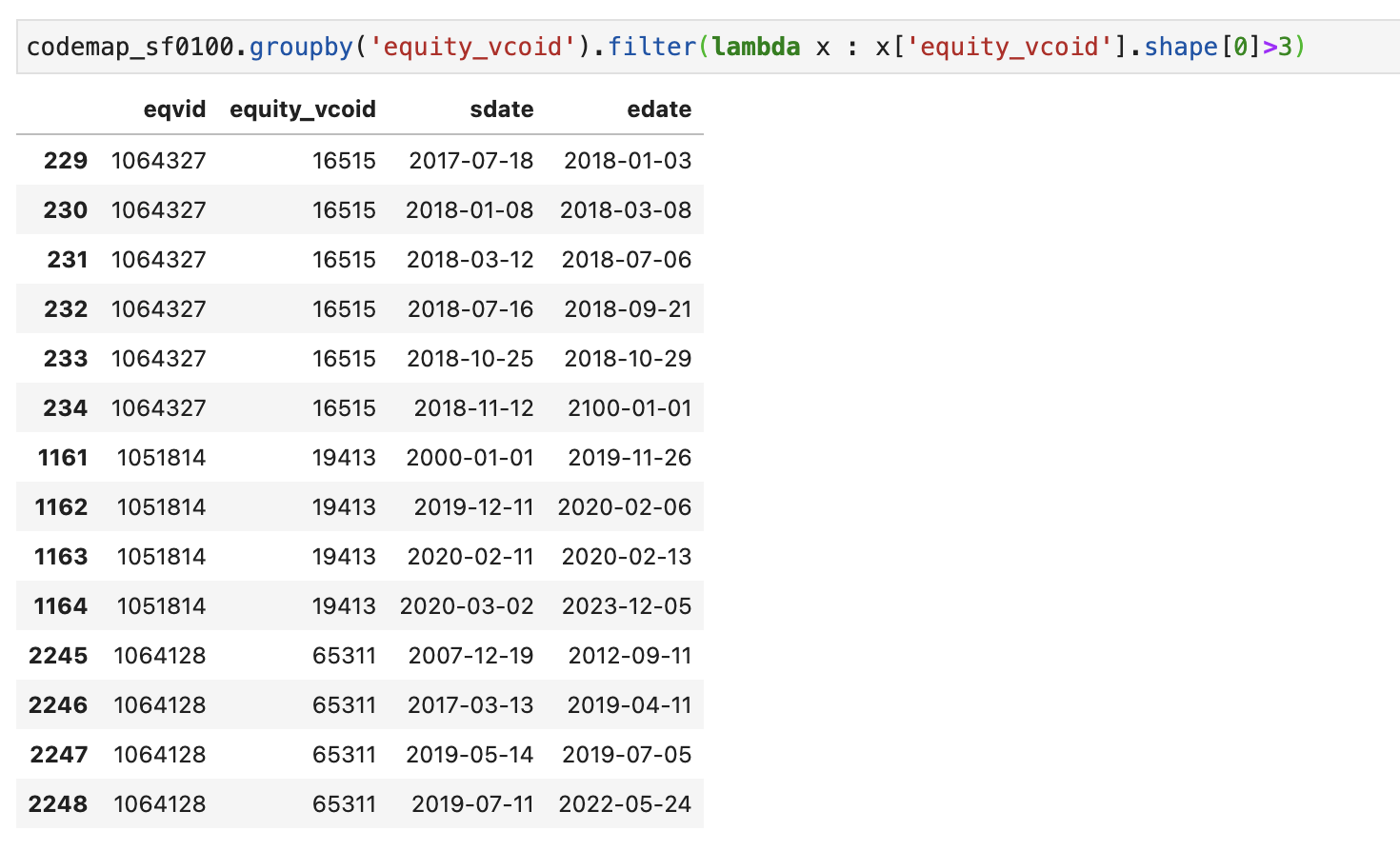
```

**Last observation carry forward in pandas**

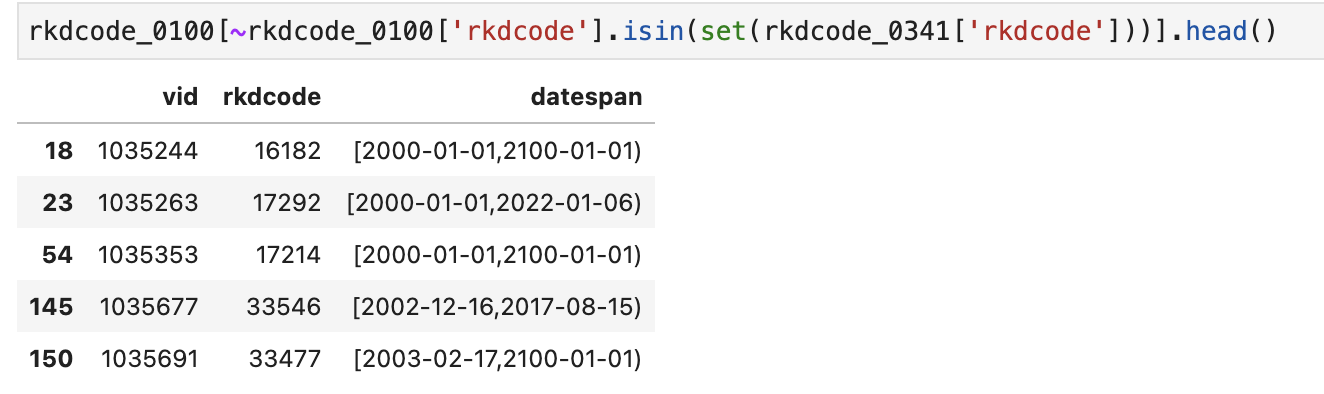
<https://www.1point3acres.com/bbs/thread-296384-1-1.html>

<https://stackoverflow.com/questions/32533722/last-value-carried-forward-in-pandas>

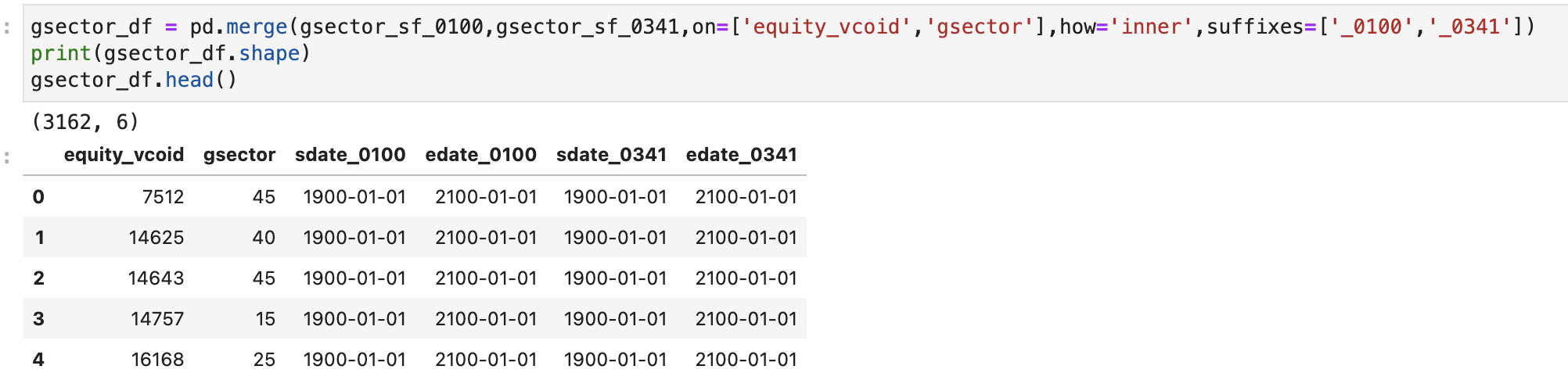
**Value\_Count > a certain number**



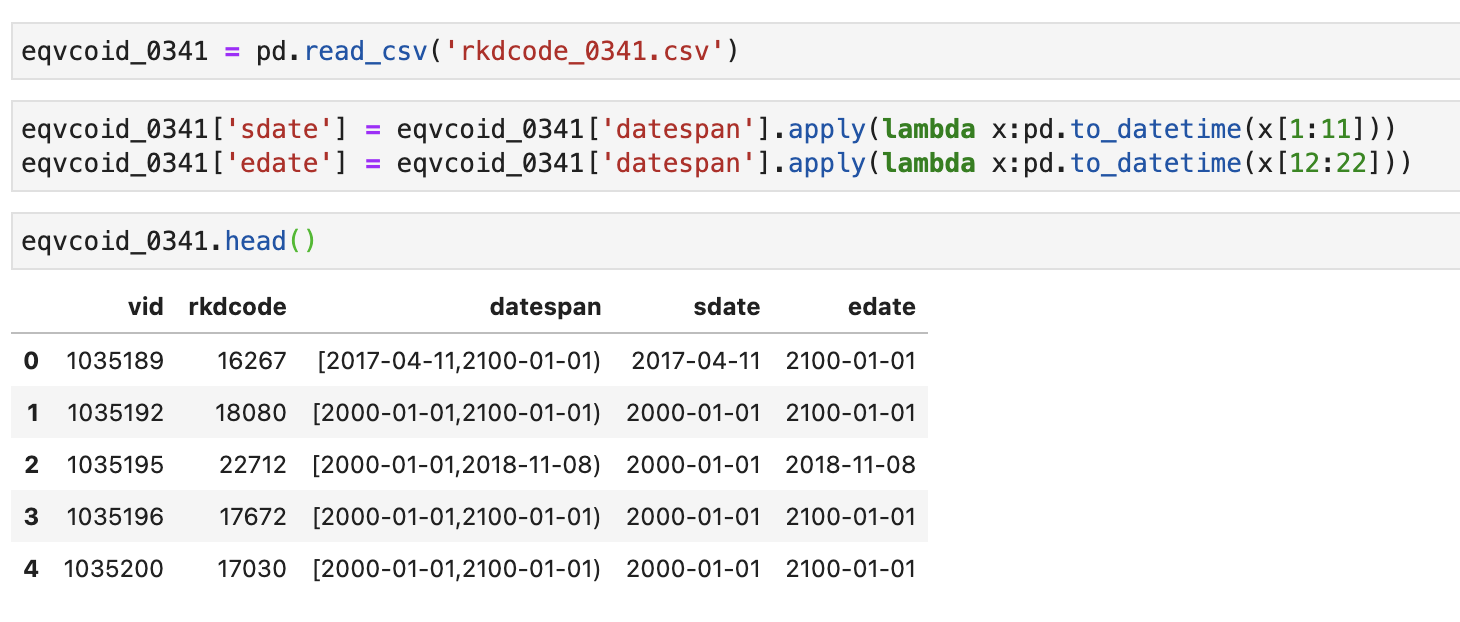
**Certain Value in one df, but also in / not in another**



**A merge example, on, how, suffixes**

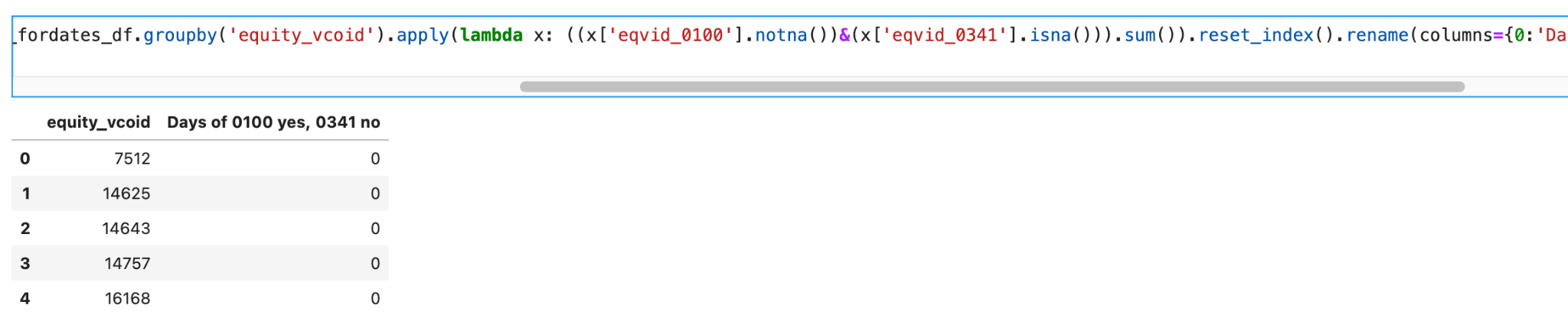
****

**Apply Lambda Function**

****

**Count the number of certain situations**

**丝滑小连招**

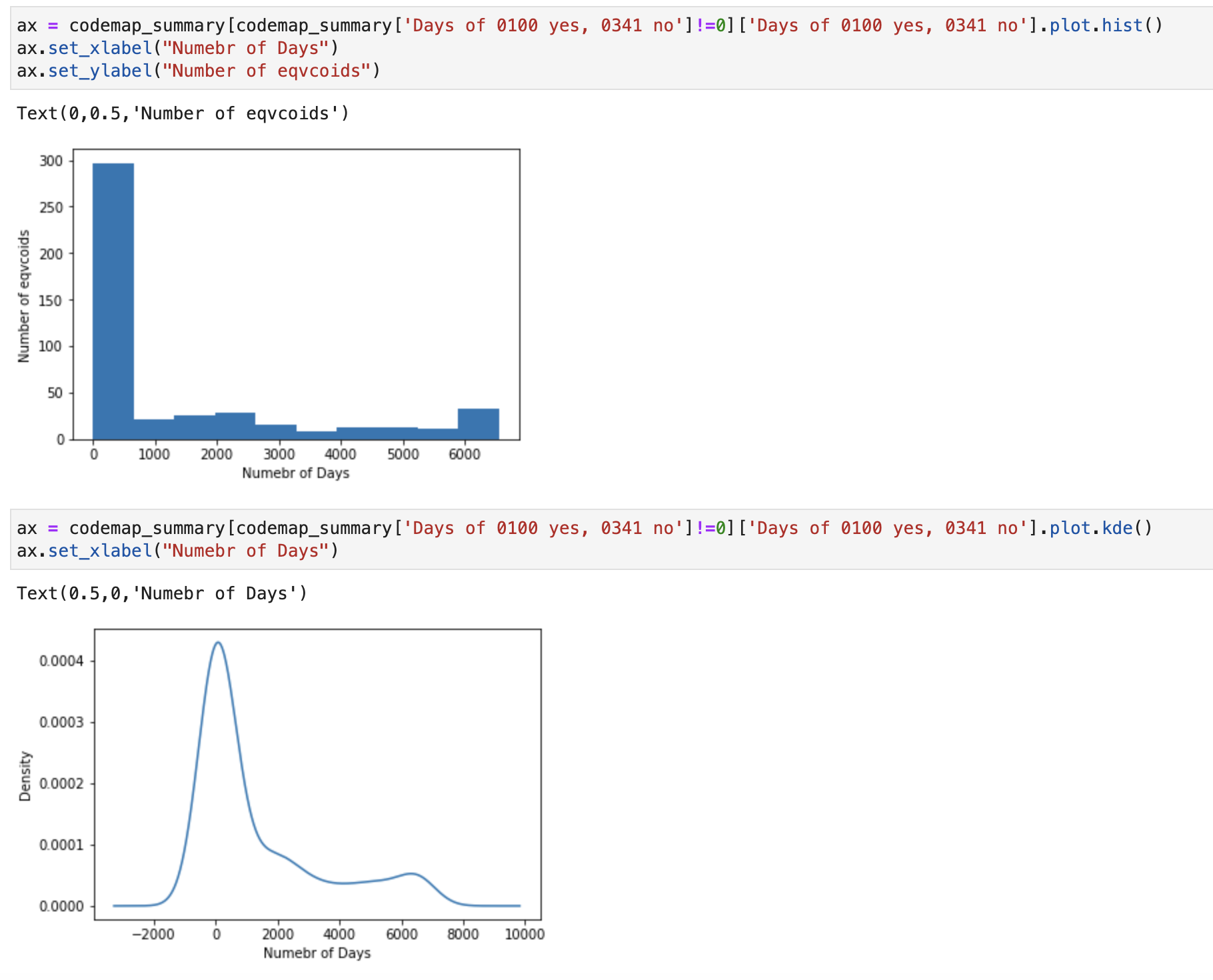
****

**reset\_index(), rename(columns={‘old’:’new’})**

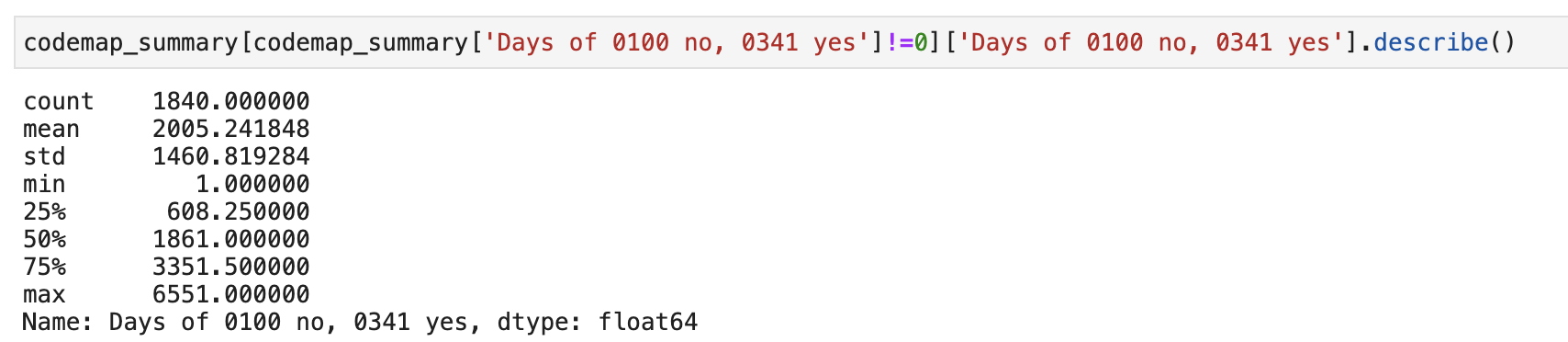
**.drop(columns=['datespan'])**

**df = df.assign(column\_B=df.column\_A)**

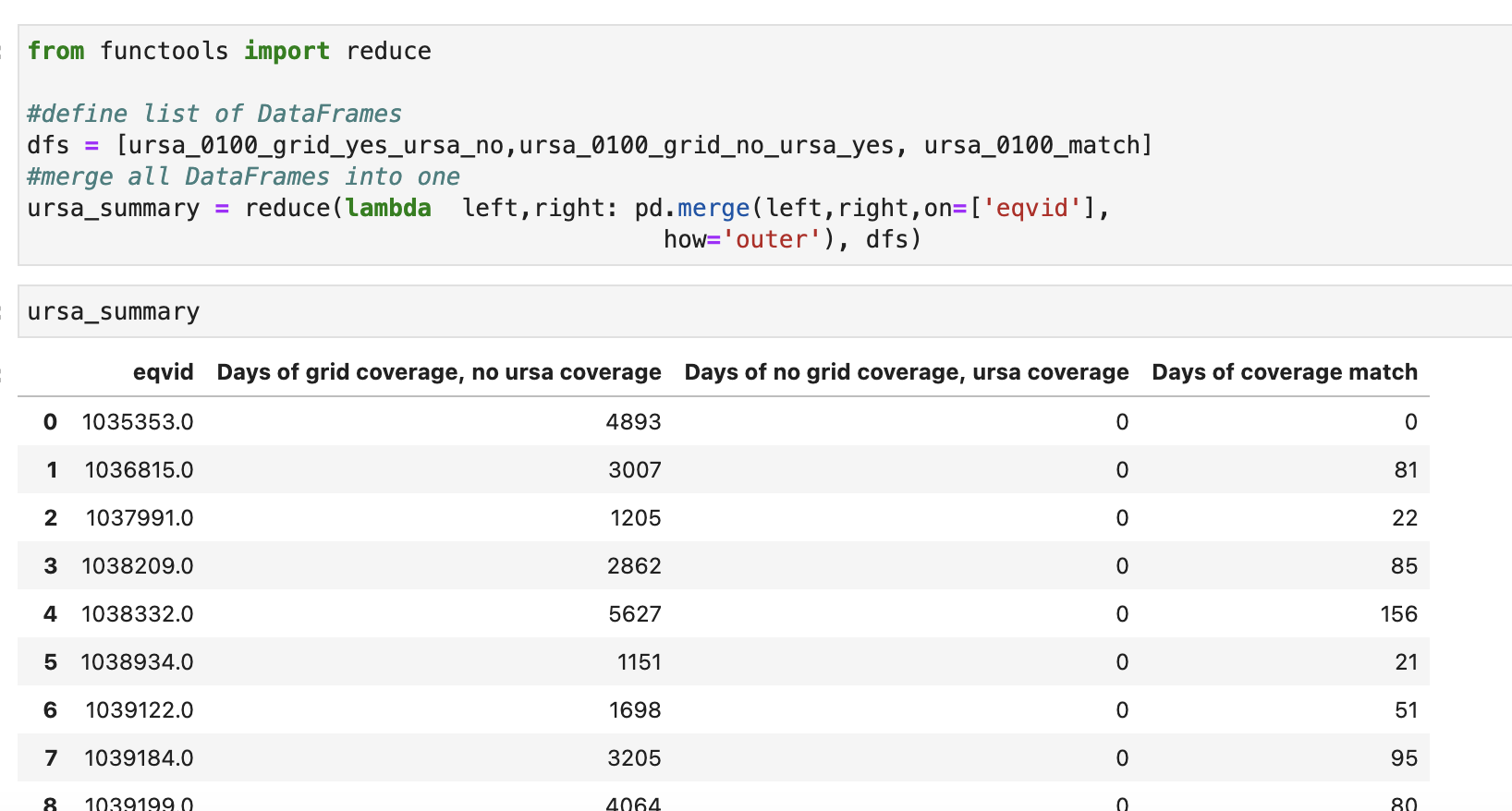
**画图，改x&y轴名字**

****

**df.describe()**

****

**merge multiple tables at once**

****