# feature

## 初始数据

### order\_products\_\_prior

order\_id product\_id add\_to\_cart\_order reordered

### order\_products\_\_train

order\_id product\_id add\_to\_cart\_order reordered

### orders

order\_id user\_id eval\_set order\_number order\_dow order\_hour\_of\_day days\_since\_prior\_order

### products

product\_id product\_name aisle\_id department\_id

## 次表

### test user

order\_id user\_id

### goods

product\_id product\_name aisle\_id department\_id department aisle

### log总表

order\_number\_rev = order\_number.max - order\_number

### order\_tbl

product\_name当前订单所有商品

t-1\_product\_name用户前一次订单所有商品

set\_diff\_products：与上次订单比新买的商品

same\_products：和上次订单买的一样的商品

### order\_aisle-department

每个订单买了哪个aisle或department，买了几件

### order\_reorderd

order\_reorderd每个订单中reorded==1的商品列表

### user\_order.csv(user\_hist)

user\_id product\_name（每个用户的所有商品） product\_id

### feature/X\_base\_t{}

is\_train

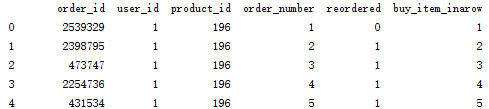
### feature/trainT-{}/label\_reordered.p

order\_id product\_id y

### feature/test/label\_reordered.p

order\_id product\_id

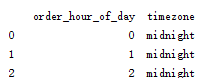
### input/mk/log\_inarow.p



### input/mk/days\_since\_last\_order/{}.p

order\_id product\_id days\_since\_last\_order\_this\_item

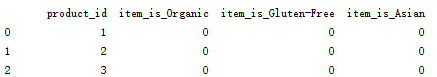
### input/mk/timezone.p





timezone

### input/mk/products\_feature.p

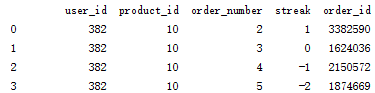


### input/mk/order\_None.p

'order\_id', 'product\_unq\_len'（个人订单累计不重复的商品数）,

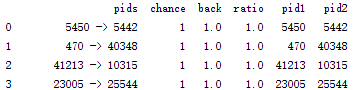
'is\_None'（本次订单购买的全是之前没买过的）

### input/mk/streak\_order-product.p



streak：对某件商品，从有购买的那个订单算起，对以后的订单，再买此商品，streak值会相对上个订单的streak增加，反之减少

### input/mk/replacement.p



（在除最后两次的订单中统计）

pid = pid1 + ‘->’ + pid2。pid1为商品号

chance：在相邻两次购物t-2，t-3中，除去共同购物的商品，从t-3中某件商品到t-2中某件的转变称为chance+1

back：从t-3中某件商品A到t-2中某件B后，再到t-1中的A，的转变称为back+1

ratio = back/chance

### order\_aisle-department\_cumsum.p



（前面还略去了很多depaertment列和aisle列）

每个department或aisle对于用户order的累加值

## user feature