# Titanic

舍弃部分feature：重复度高或包含很多NAN的

创造部分feature：如把兄弟姐妹和父母，创造一个family；把连续的，划分出几个分段，如年龄；从名字中提取出，头衔，这个feature

对没有空NAN值的数据进行比例分析，如

|  | **Pclass** | **Survived** |
| --- | --- | --- |
| **0** | 1 | 0.629630 |
| **1** | 2 | 0.472826 |
| **2** | 3 | 0.242363 |

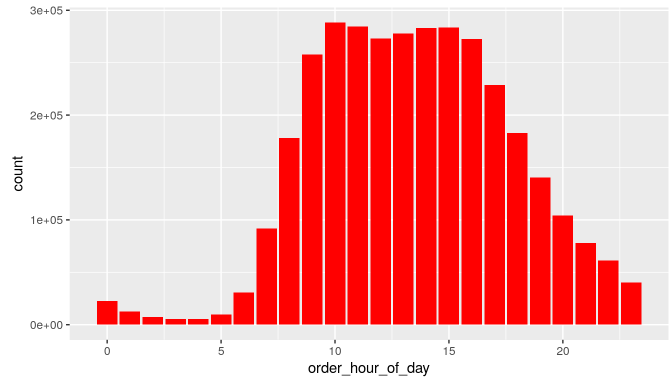
# Instacart Market Basket Analysis

## 初步分析

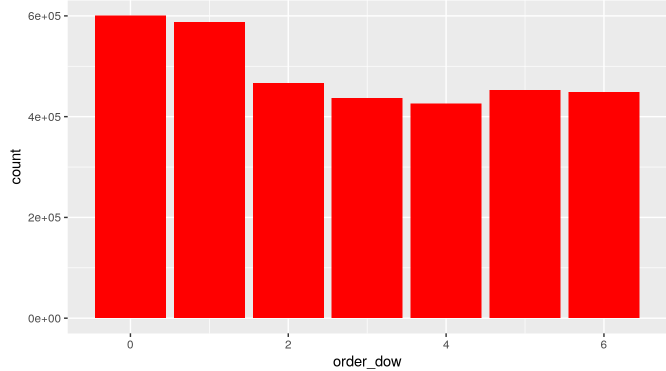
、、、、、、

商品销售额、商品再购率、department销量、department再构率、aisle销量、aisle再购率、用户总订单数、是否为有机产品organic

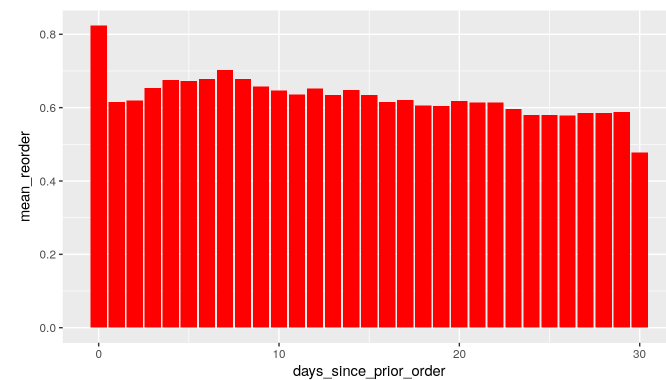
### order\_hour\_of\_day



### order\_dow

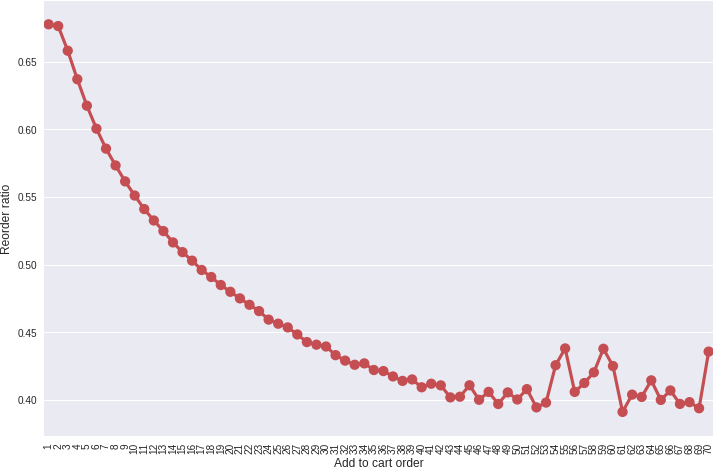


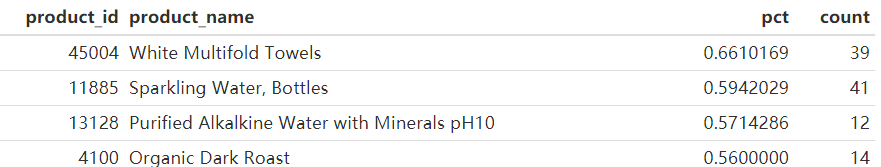
### days\_since\_prior\_order



### order\_number

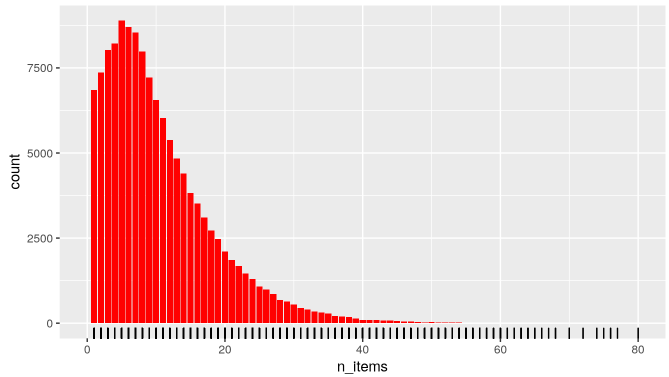
### add\_to\_cart\_order





### item\_number\_in\_a\_order

How many items do people buy

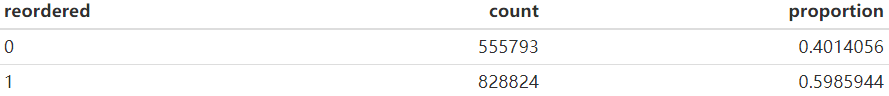


### 商品总销量

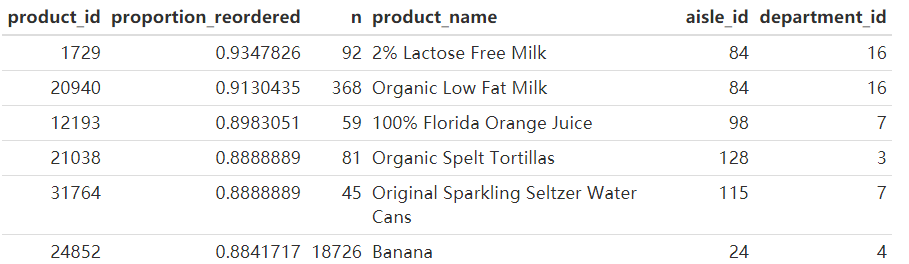


Bestsellers

### 商品再购率

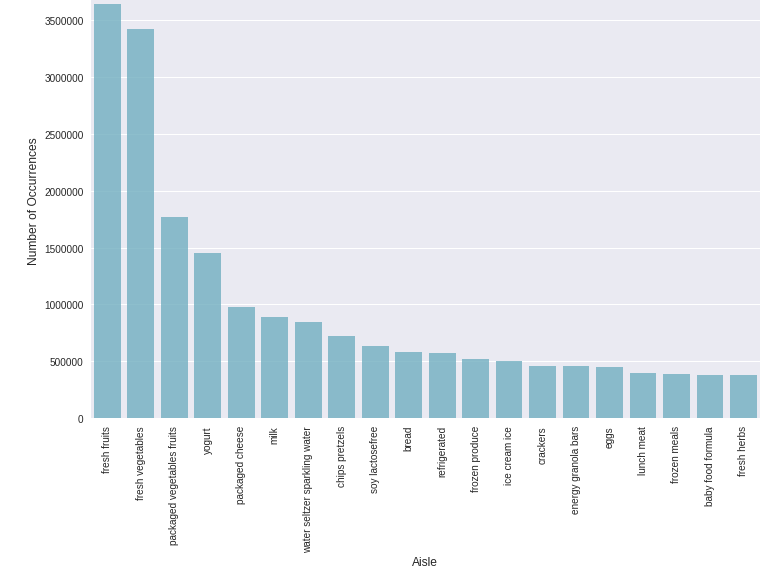


59% of the ordered items are reorders.

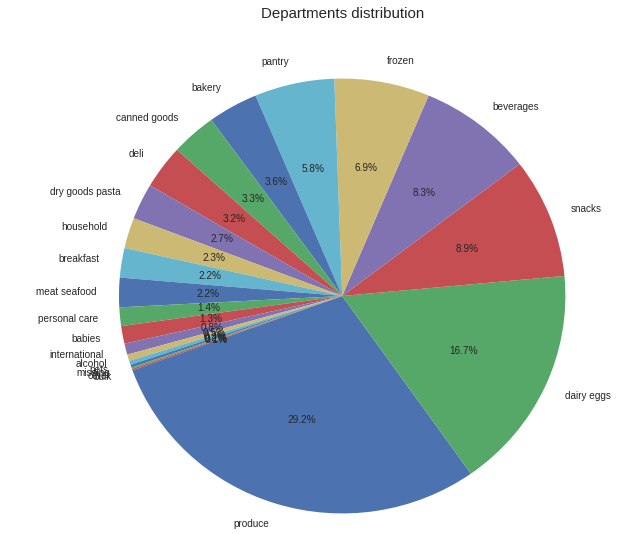


Most often reordered

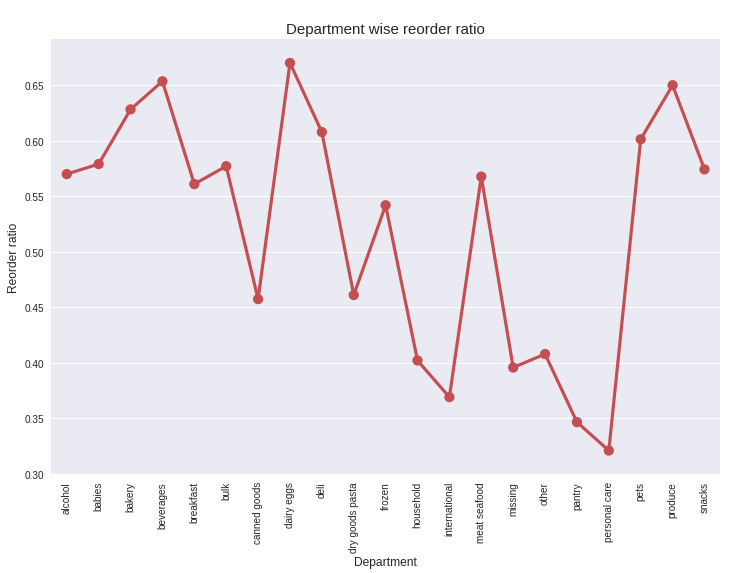
### aisle销量



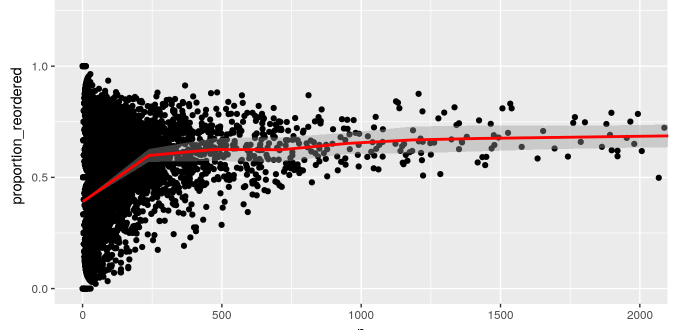
### department销量



### department再购率



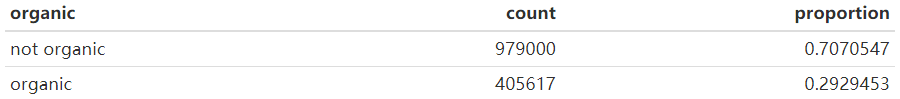
### 订单数

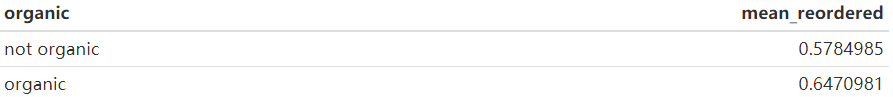


Association between number of orders and probability of reordering

订单数越多，重买率越高

### 有机产品





### 是否买永远买同样的商品

3,487人永远在买同样的商品

## user相关feature

用户平均days\_since\_prior\_order值，用户订单总量，用户总购买商品数，用户平均每个订单的商品数，用户不重复的总商品数

## order相关feature

该order在几点下单，和上次订单相距时长，相距时长比（=相距时长/用户平均相距时长）

## product相关feature

aisle\_id、department\_id、商品销量、商品再购量、商品再购率

## user\_X\_product相关feature

该商品在此用户的销量，销量占用户总商品比，last\_order\_id，平均加入购物车的顺序，再购率，和最后一次购买此商品相隔订单数、购买时小时数相隔时长