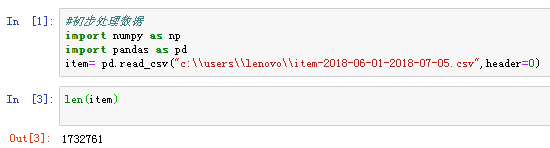
# 第一次超市订购单明细报告分析

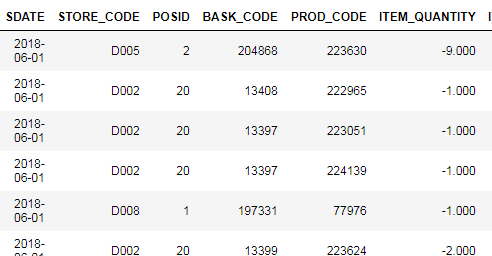
## 超市订购单明细数据初步处理

利用python导入原表item，观察到共有订购交易记录1732761项，超市订够指标列分别表示交易时间、商店编码、POS机编码、商品栏编码、商品编码、购买商品数量、购买实际金额、正常购买金额、折扣类型和折扣金额。本报告首先根据指标列对存在不合理的交易记录进行初步处理，并对存在问题的交易记录进行更正及删除。



1. 交易数量问题

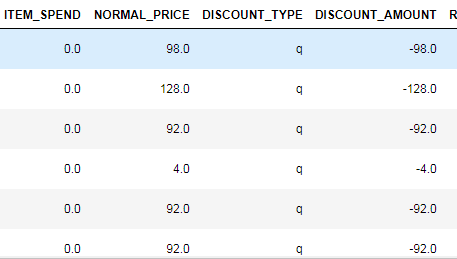
如下图所示。超市交易明细中记录2022条的交易数目存在问题，问题在于交易数量为负值。按照实际交易情况分析，交易数目应为非负数，因此本报告将2022条明细记录中的交易数量更正为正数，更新item表。



1. 实际交易金额问题

各商店实际交易金额并存在与以上与交易数量相同的问题，即实际交易金额存在负数的情况为2020条明细，观察处理数据发现交易数量与实际交易金额均为负值可能导致异常实际折扣数值异常，因此本次在python中对存在负值问题的数据进行了绝对值处理，将其同样更正为正值。

如下图所示，通过数据显示在交易记录明细当中有636条交易记录的实际交易金额为0元，且折扣类型都为Q。超市在6月1日—7月4日期间对部分商品进行全折销售。



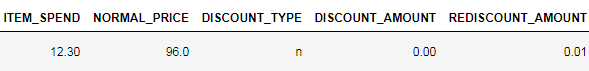
3.折扣金额问题

由于交易数量以及实际交易金额出现问题，本次对折扣金额也同样做出对比计算来验证再计算折扣金额和实际各商店记录折扣金额是否存在出入。本次通过简单计算编写算式表达式重新计算折扣金额，表达式为实际交易金额-交易数量\*正常交易金额，且按照大多数商店计数规则保留2位小数。

对前步骤更改过后，结果对比发现在1732761项交易记录明细当中有6524条记录存在不对等的情况。经过初步分析导致DISCOUNT\_AMOUNT和RDISCOUNT\_AMOUNT不对等的原因有三类：

1. 各商店计算折扣价格计数问题

如下图所示，可以明显地看到DISCOUNT\_AMOUNT和RDISCOUNT\_AMOUNT这两列不相等，原因在于编码为D008的 商店对17893栏113167商品计算折扣价格时保留一位小数点引发的差异。



1. 商店记录折扣价格时，记录方式不统一

如下图，原数据交易明细中DISCOUNT\_AMOUNT存在两种记录方式，例如折扣价格-40和40元，因此在数据进入下一步分析应统一记录方式。



1. 交易明细记录错误

例如下图为2018年7月5日商店A018售卖215524栏29318商品的交易记录，其折扣金额为-1440元，明显出现记录失误，对于此类情况应予以删除此类记录的处理。



## 超市交易明细指标列分析

1.交易记录时间范围

原交易明细共有1732761项，按照时间对每条记录进行排序发现交易明细记录了在2018年6月1日-2018年7月5日的35天各商店交易情况。其中2018年6月1-2018年7月5日的交易按天交易明细计数统计图，如下所示：

如图中显示，2018年6月1日-7月5日商店交易数量呈现上下波动的情况，日交易高峰数量在2018年6月17日左右，17日后直到7月5日交易数量明显低于17日之前的交易数量。由此看来，超市每日交易次数下滑。

2.商店编号情况

编码为A类的商店有15家，02、03、05、11、14、15、20商店的交易记录没有记录在商店内。编码为D类的商店共9家，其中01商店的交易记录无法得知。商店D003及A001分别为销售第一、二的热门商店、最为冷清的商店为D003。

|  |  |
| --- | --- |
| 商店编码 | 交易明细数量 |
| A001 | 193822 |
| A004 | 46005 |
| A006 | 77906 |
| A007 | 49326 |
| A008 | 106171 |
| A009 | 56326 |
| A010 | 49291 |
| A012 | 63749 |
| A013 | 81522 |
| A016 | 56266 |
| A017 | 28789 |
| A018 | 46569 |
| A019 | 51289 |
| A021 | 62889 |
| A022 | 44130 |
| D002 | 242492 |
| D003 | 22249 |
| D004 | 76548 |
| D005 | 50408 |
| D006 | 70071 |
| D007 | 117804 |
| D008 | 59282 |
| D009 | 45032 |
| D010 | 34825 |

3.POS机情况

POS机情况图描述了各商店对不同编号的POS机使用情况，所有交易明细当中记录了1-10、13、19、20号POS机的交易情况，1号和2号POS机的使用次数多于其他机器。

4.商品栏情况

统计所有商店有281211个不同的商品栏。