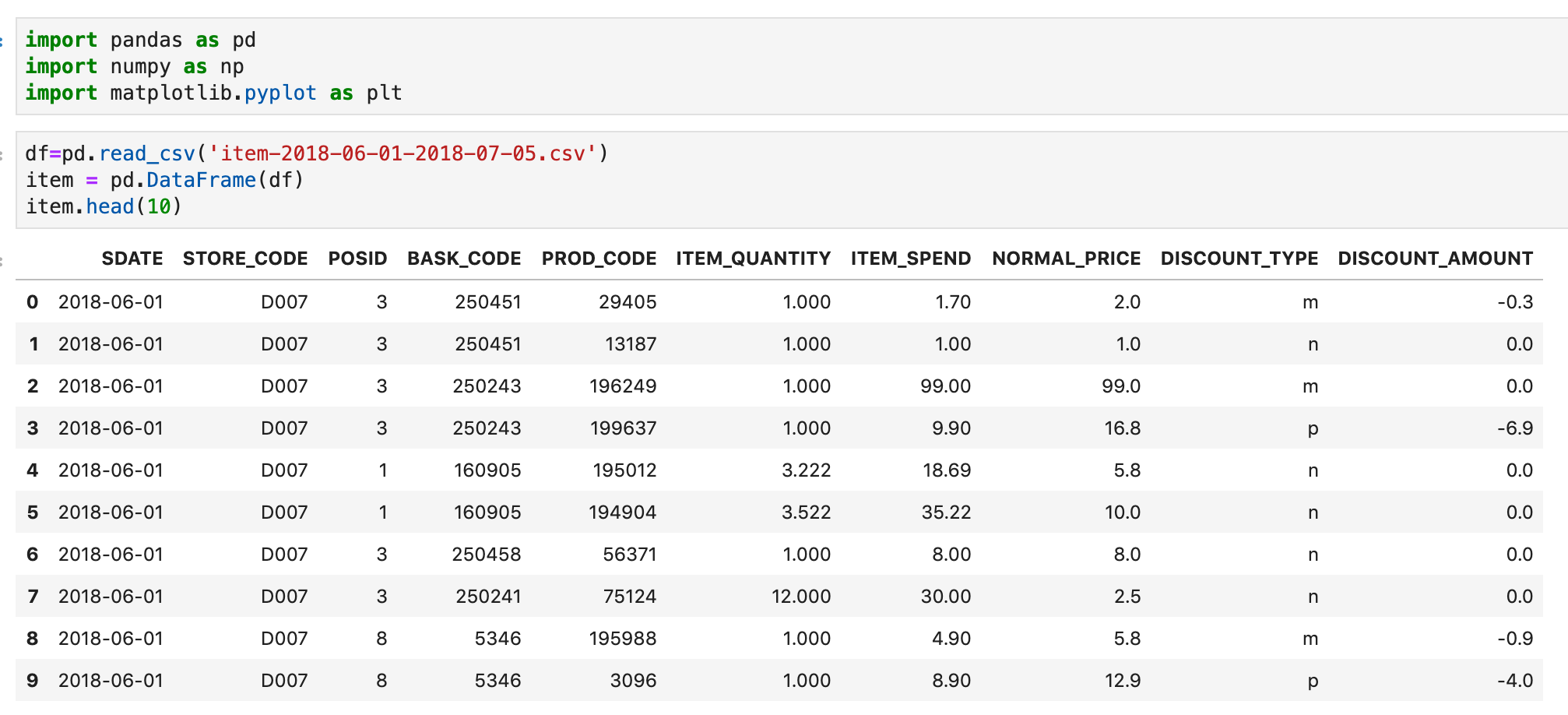
数据分析大作业报告

1. 样本集

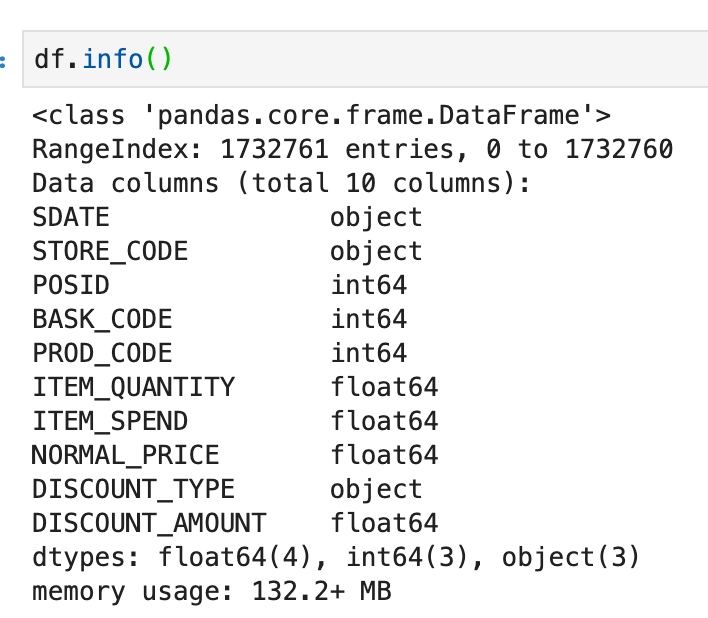
数据集来源于（item-2018-06-01-2018-07-05.csv、Basket-2018-06-01-2018-07-05.csv）。

1. 数据读取与预处理

2.1首先导入库和展示数据。

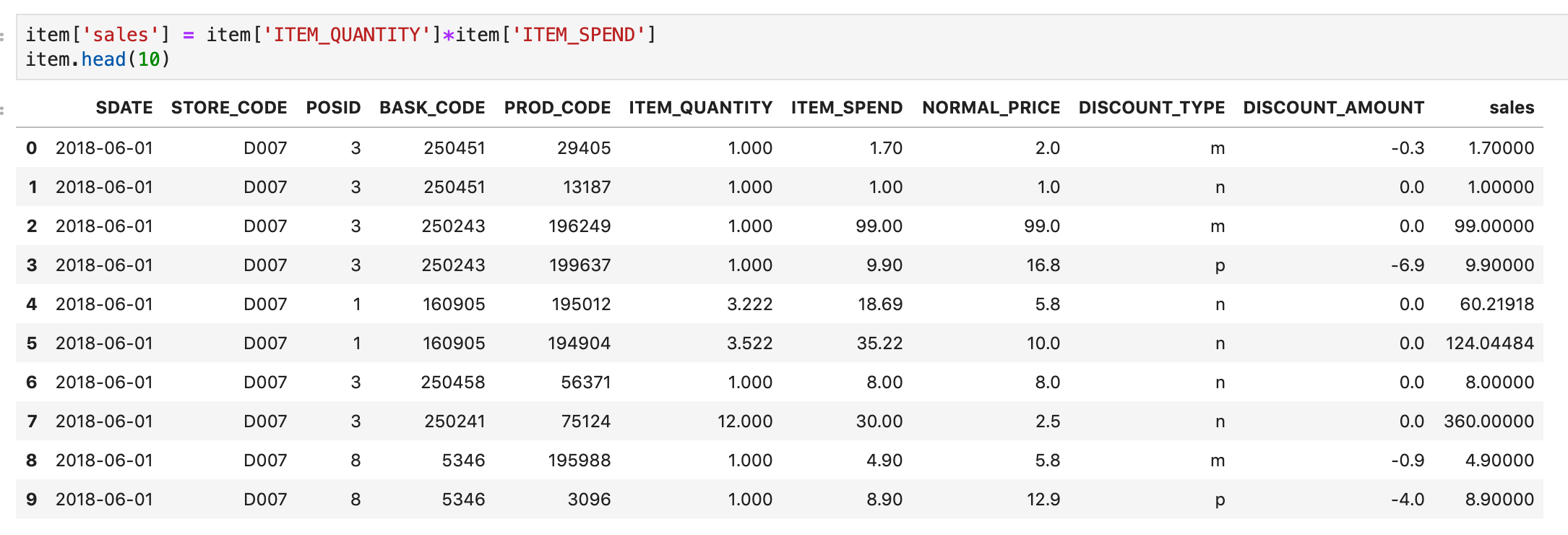


2.2查看索引、数据类型和内存信息。



可以看到我们需要用于计数的特征值已经是浮点数类型，不需要再更改。并且数据只有产品数量和单价，没有销售金额，要根据产品数量和单价建立一个销售金额的新字段。

2.3新建一个实际销售金额的列，即用产品数量乘以折后价格。



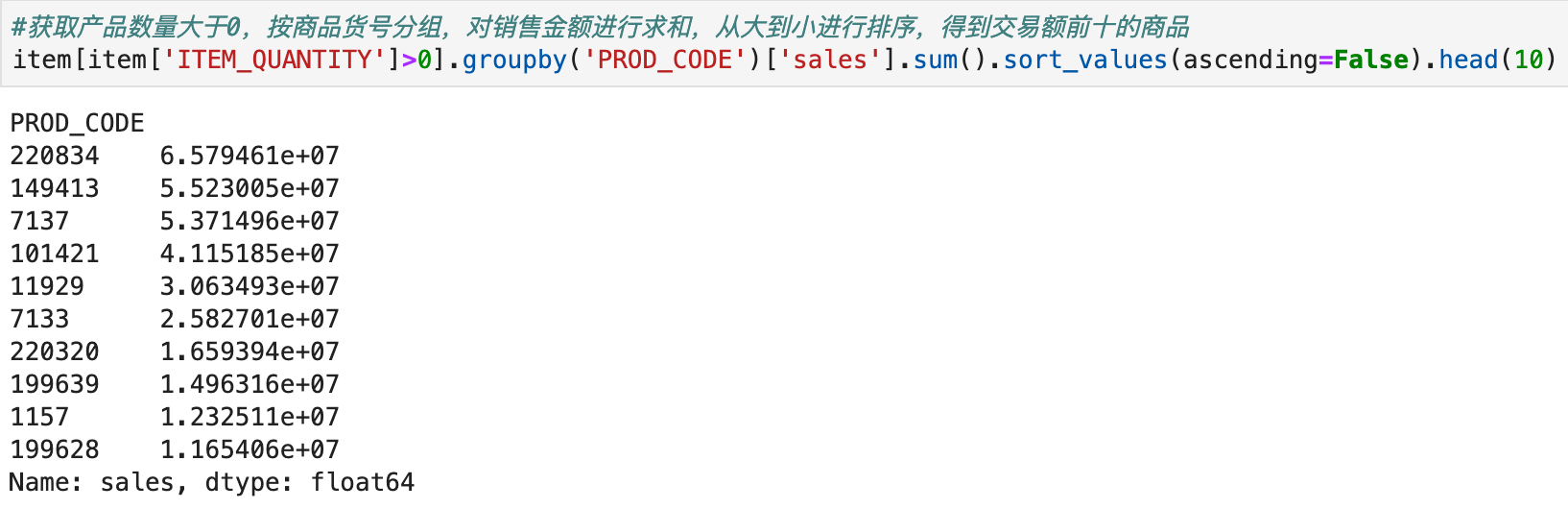
1. 数据分析

3.1获取产品数量大于0，按商品货号分组，对商品数量进行求和，从大到小进行排序，得到销售量前十的商品。



通过结果我们可以得到2018年6月1日到7月5日之间销量最高的商品编号分别为149413、86772、220834、7133、61623、124556、205540、1152、194690、124554。前几位商品有可能是该超市销售的主要对象。商家可以通过这个数据对商品在顾客心中的喜爱度做一个基本的衡量。

3.2获取产品数量大于0，按商品货号分组，对销售金额进行求和，从大到小进行排序，得到交易额前十的商品。



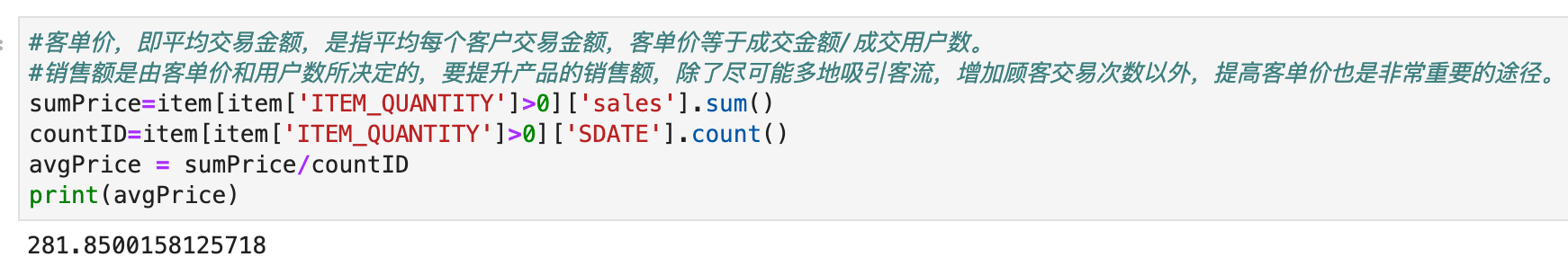
通过排序得到的销售额最高的数据，可以避免片面的看待销量最高但单价很低，也就是性价比比较低的商品，而总销售额和销售量对比起来分析可以看出销量和售价都相对优良的高性价比商品。所以除了销量高的几个商品以外，交易额跻身前十的商品也应该重点关注。

3.3获取产品数量大于0，按日期对销售额进行排序，得到日销量的排名。



可以看出销量最高的是6月17日，其次是6月18日和6月16日。大部分销售额集中在六月中旬期间，即大部分工薪阶层发工资的期间，可估计在月中商店的生意最为火爆，可以相应的推出活动刺激更多的消费，也可以选择在销售额较为低的月初和月末进行一些促销推动销售额的增长。

3.4客单价即平均交易金额，是指平均每个客户交易金额，客单价等于成交金额/成交用户数。



销售额是由客单价和用户数所决定的，要提升产品的销售额，除了尽可能多地吸引客流，增加顾客交易次数以外，提高客单价也是非常重要的途径。