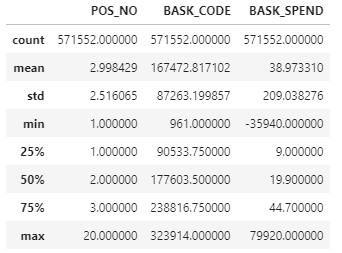
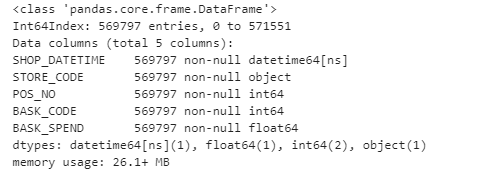
数据分析

1. **描述性分析**
   1. **数据清洗**

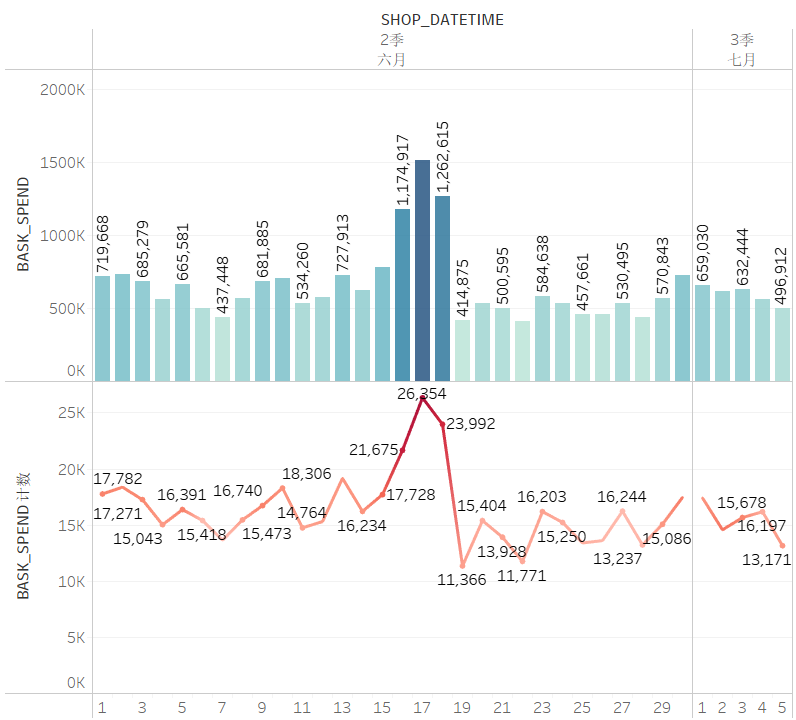
查看数据描述信息，如下图1所示，可以看出BASK\_SPEND在最小值出现了负值情况，筛出BASH\_SPEND小于0的数据，筛选后的数据描述如图2 所示，数据条数为569797条。



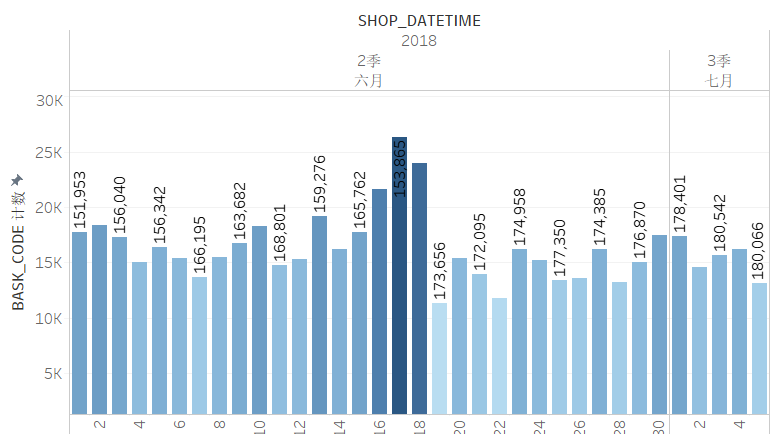


* 1. **数据分布情况**

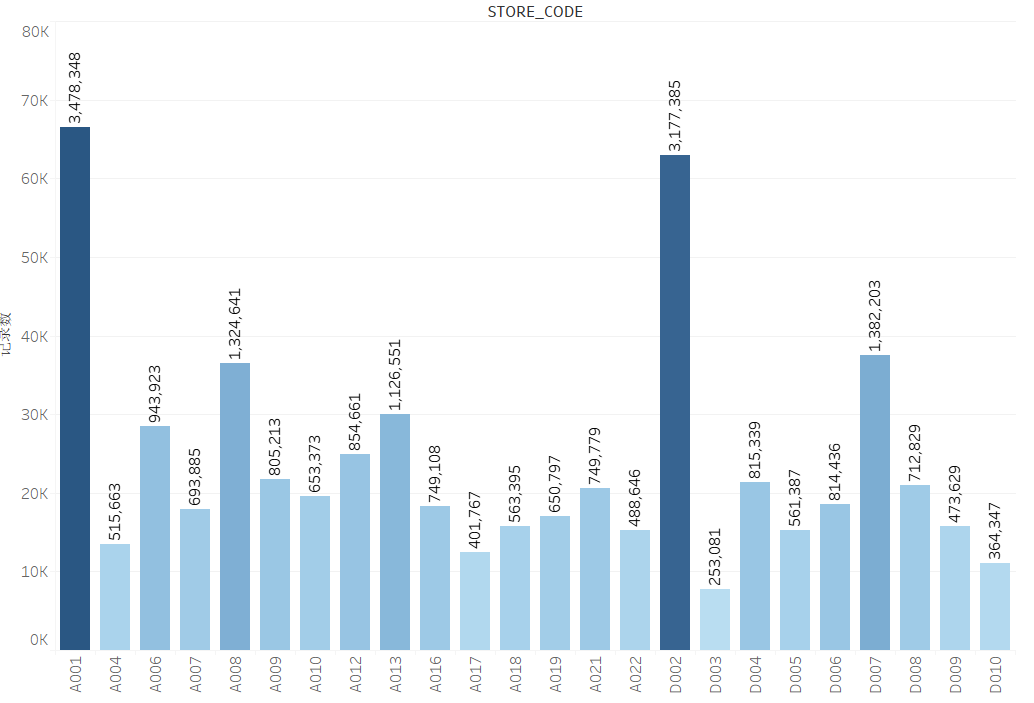
**每日消费数额与消费笔数分布：**



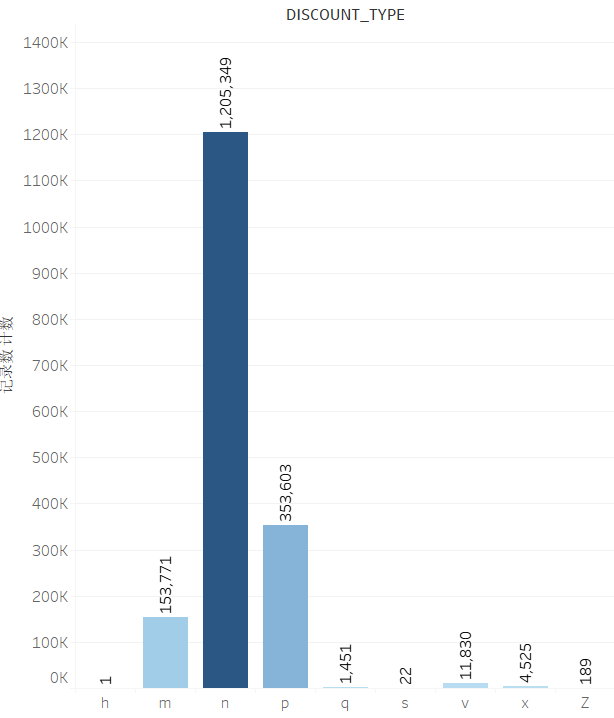
**消费人数分布情况：**



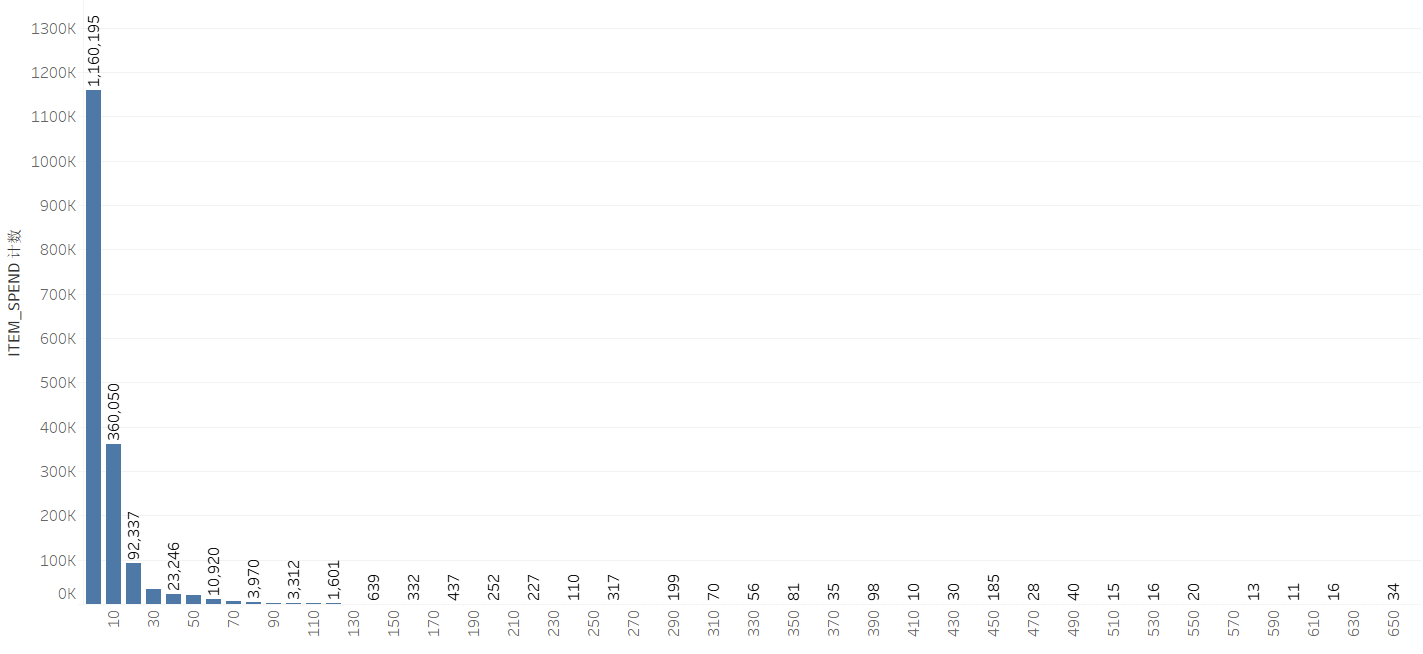
**产品销售情况分布图：**



**折扣情况商品数量分布图：**



**商品价格分组分布情况：**



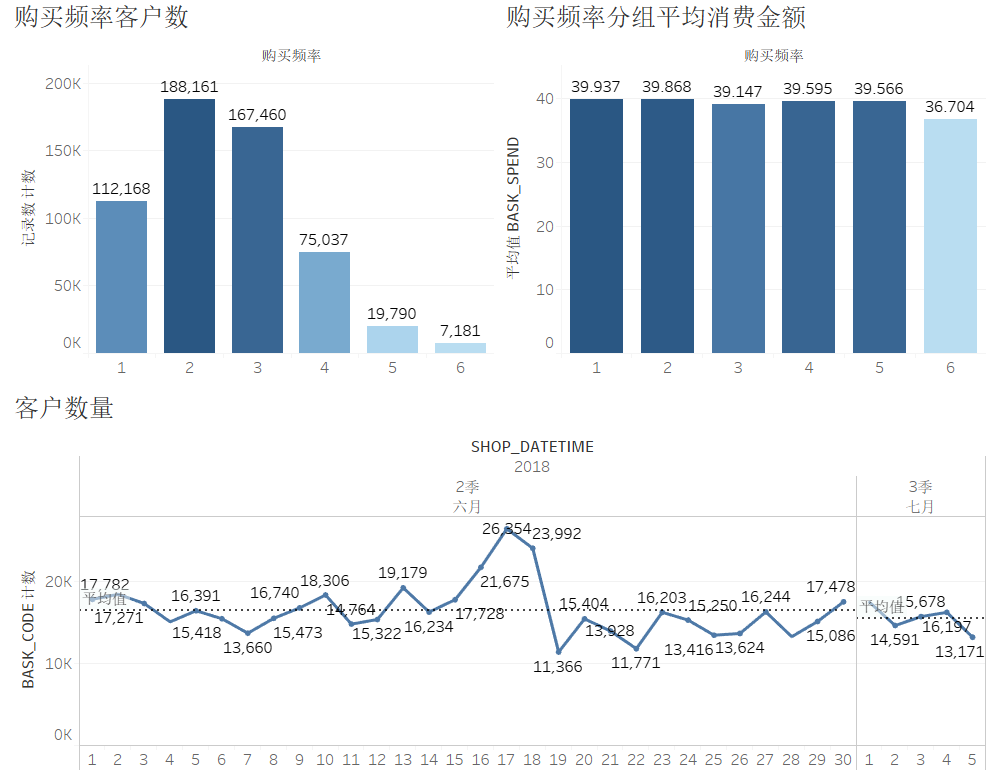
### 二．客户相关数据分析

**2.1 客户购买频率等分析**

如下图所示，通过计算用户的购买频率，形成购买频率客户数图，经过分析用户购买频率分为1-6次，在统计期间购买2-3次的人数最多，分别为188802人次和167914人次。

通过购买频率分组平均消费金额的图像，可以看出购买频率为1-5次的客户的平均消费金额差距不大，都为39元左右。消费频率最高的6次的用户，平均消费额与前几个分组频率有着明显区别，为36.499元，可能与用户的购买频率较高有关联。

通过客户购买人数的日期分布来看，六月平均购买人数为16431人，七月为15468人，这与七月这五天为周一至周五工作日有关，并不能说明人流量出现明显下滑，其中六月16、17、18三天的流量情况异常突出，星期分别为星期六、星期日、星期一，可能又相关促销行动举行，可在商品分析时，进一步分析。但19日也就是周二的时候出现了较大的人流量下滑，活动过后的人流量下滑问题可进一步思考，提出对应策略分析解决。



**2.2 RFM模型**

RFM模型是一个简单的根据客户的活跃程度和交易金额贡献所做的分类。因为操作简单，所以较为常用。

**近度R：**R代表客户最近的活跃时间距离数据采集点的时间距离，R越大，表示客户越久未发生交易，R越小，表示客户越近有交易发生。R越大则客户越可能会“沉睡”，流失的可能性越大。在这部分客户中，可能有些优质客户，值得公司通过一定的营销手段进行激活。

**频度F：**F代表客户过去某段时间内的活跃频率。F越大，则表示客户同本公司的交易越频繁，不仅仅给公司带来人气，也带来稳定的现金流，是非常忠诚的客户；F越小，则表示客户不够活跃，且可能是竞争对手的常客。针对F较小、且消费额较大的客户，需要推出一定的竞争策略，将这批客户从竞争对手中争取过来。

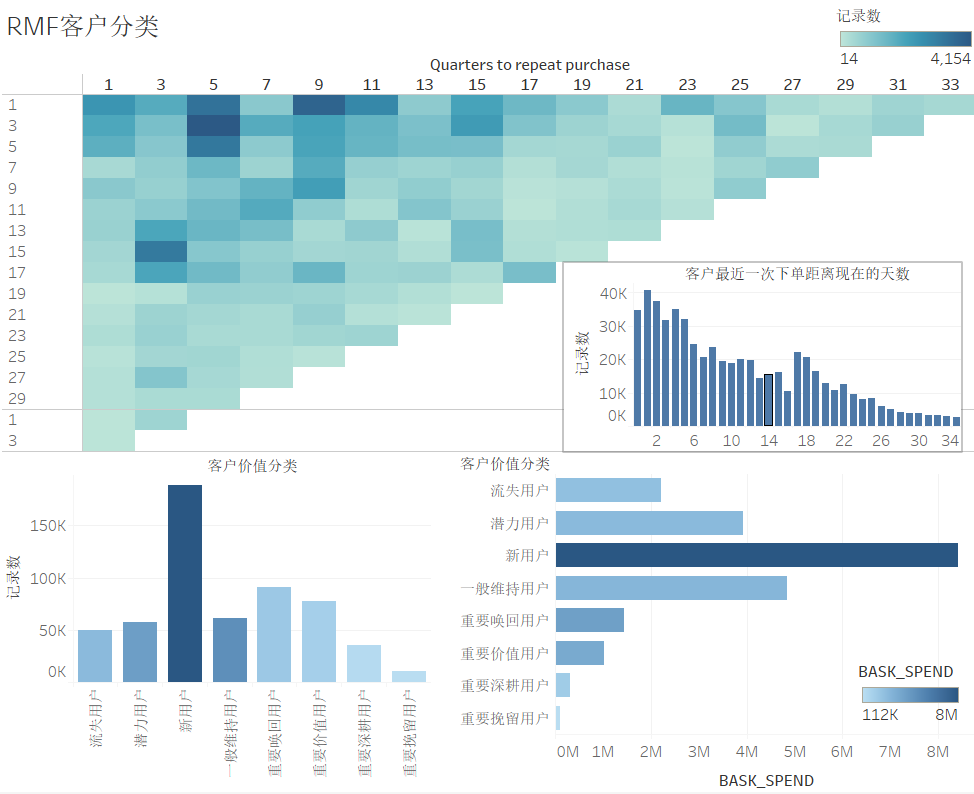
**额度M：**表示客户每次消费金额的多少，可以用最近一次消费金额，也可以用过去的平均消费金额，根据分析的目的不同，可以有不同的标识方法。一般来讲，单次交易金额较大的客户，支付能力强，价格敏感度低，是较为优质的客户，而每次交易金额很小的客户，可能在支付能力和支付意愿上较低。当然，也不是绝对的。

通过RFM分析将客户群体划分成一般保持客户、一般发展客户、一般价值客户、一般挽留客户、重要保持客户、重要发展客户、重要价值客户、要挽留客户等八个级别。



如下图所示，通过购买间隔天数热力图可以看出，自第一天购买日期，间隔五到九天复购的人数最多。而根据最后一天下单距离表格统计日期，可以看出大多用户在前1-5天有过购买行为。

从客户价值分类来看，新用户占比较大，属于最近有过购买记录但是累计单数累计消费金额较少的人群，但可能是因为统计时间的问题，造成用户累计单量未被完全统计，而用户分类出现偏差，新用户人群占比提升。



**2.3 工作日划分分析**

如下图所示，用树状图表现星期一至星期日的销售情况，销售量表示图形大小，颜色表示销售额，颜色越深表示销售额越大。根据图形可以看出周六周日销售量和销售额相当，平均单价也很接近分别为42.801和42.918，领先其他工作日销售情况。工作日中，紧邻周末的周一和周五销售情况最好，整个星期的消费情况呈现一个凹的造型。

按日期分析销售情况可以看出六月16、17、18三天的销售情况异常突出，星期分别为星期六、星期日、星期一，可能又相关促销行动举行，可在商品分析时，进一步分析。

