淘宝用户行为分析案例——天池数据集User Behavior Data from Taobao

说明

本次数据分析基于阿里云天池数据集（[用户行为数据集](https://tianchi.aliyun.com/dataset/dataDetail?dataId=649)），使用转化漏斗，AARRR模型，对常见电商分析指标，包括转化率，PV,UV,留存率等进行分析，分析过程中使用MySQL以及Navicat进行数据预览与清洗处理，使用Excel进行数据可视化。

一、数据集与分析目的

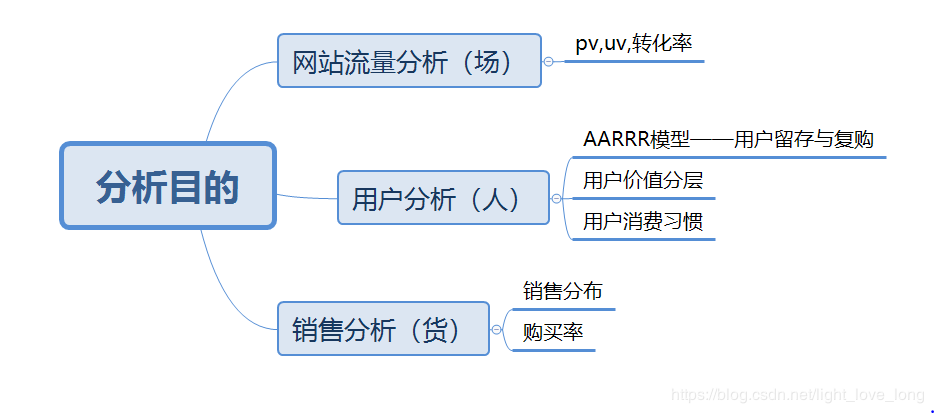
1、数据集



该数据集记录用户在淘宝网站浏览商品产生的行为信息。由于数据集过大，选取其中105万条数据进行分析。

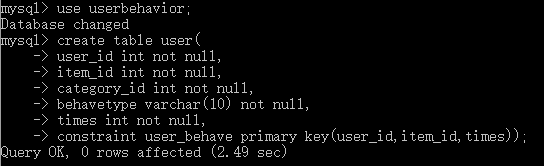
1. 分析目的
2. 了解网站流量情况；
3. 了解该阶段网站用户粘性以及用户行为习惯；
4. 了解网站商品销售情况；

分析逻辑如下：

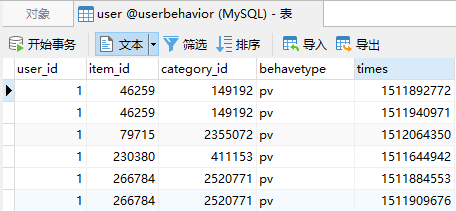


1. 数据处理
2. 数据预览

MySQL建表，列名重命名：

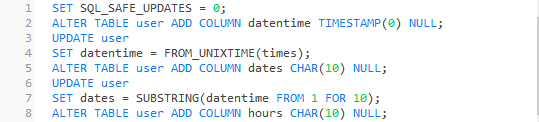


使用Navicat导入数据：



1. 数据处理
2. 日期处理

将行为发生时间转为datetime类型，获取其发生日期及时间：

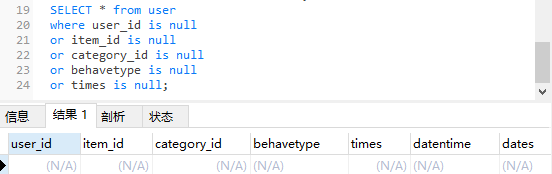




1. 选取数据：选取2017-11-25至2017-12-04之间的数据进行分析

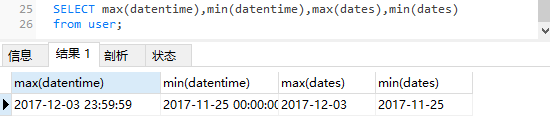
66c7a97105dc492115e88de9a51b85d

1. 查看缺失值



数据集中不存在缺失值。

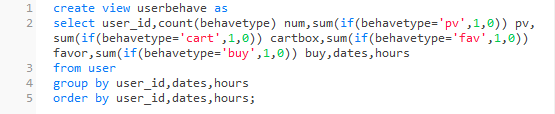
1. 查看异常值：主要查看是否存在异常时间

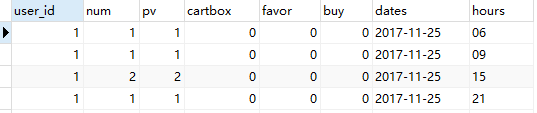


日期不存在异常值。

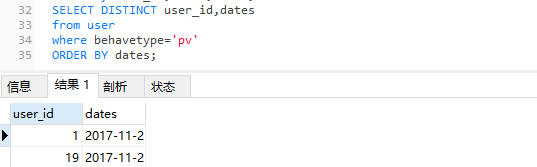
1. 数据提取与分析
2. 网站流量分析

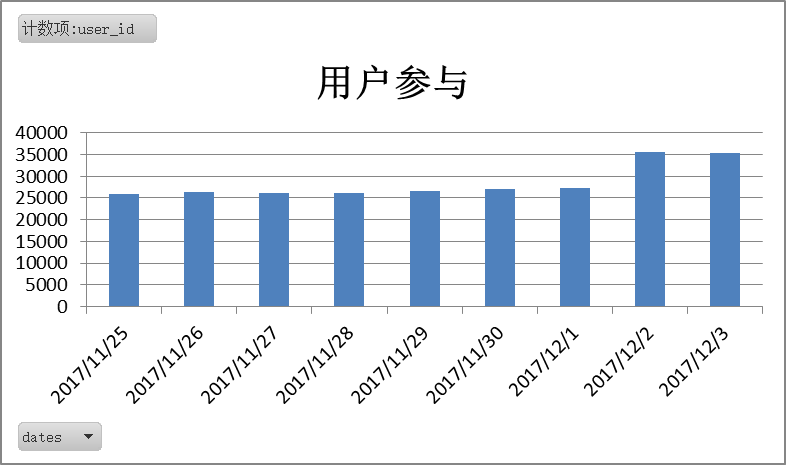
创建用户行为视图：





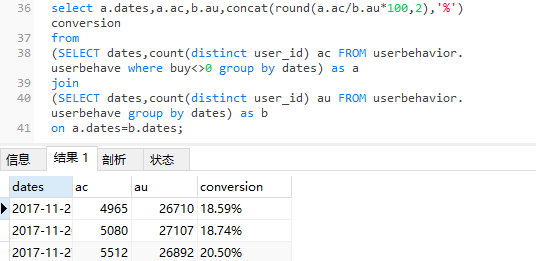
1. 用户参与度（PV）分析

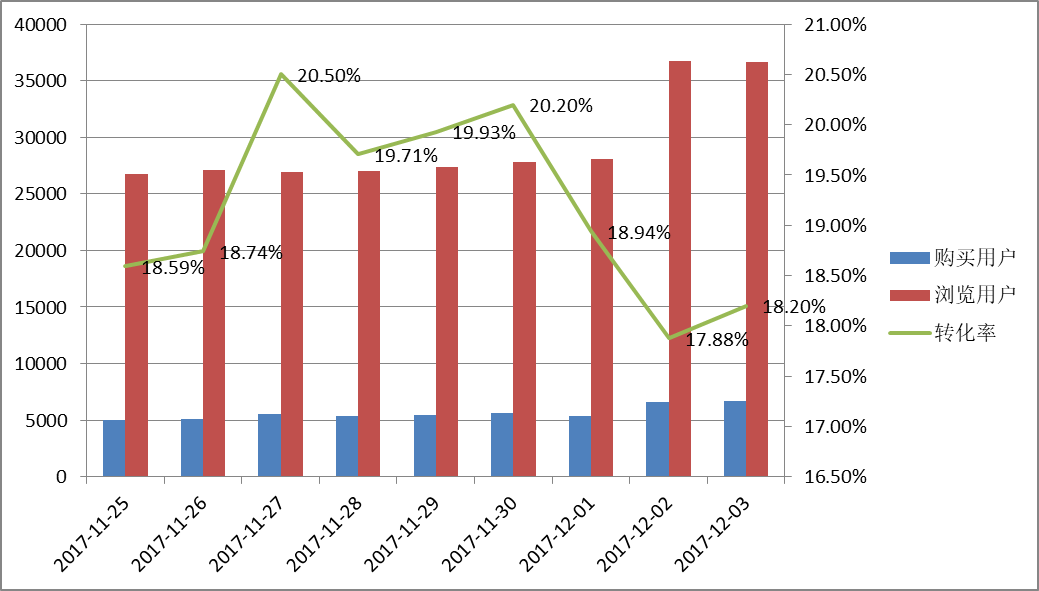




分析：11月25日至12月1日用户数变化幅度较小，范围在25000-30000之间，从12月2日开始有较大幅度增长，对比11月25，26（周末）有较大提升，可能与周末的双十二营销活动有关。

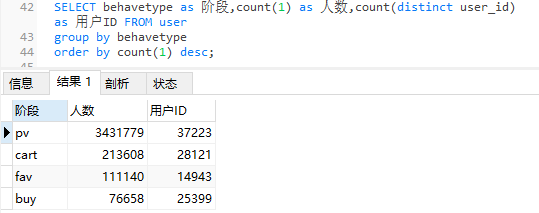
1. 网站购买转化率（UV）：（支付转化率=支付访客数/总访客数）



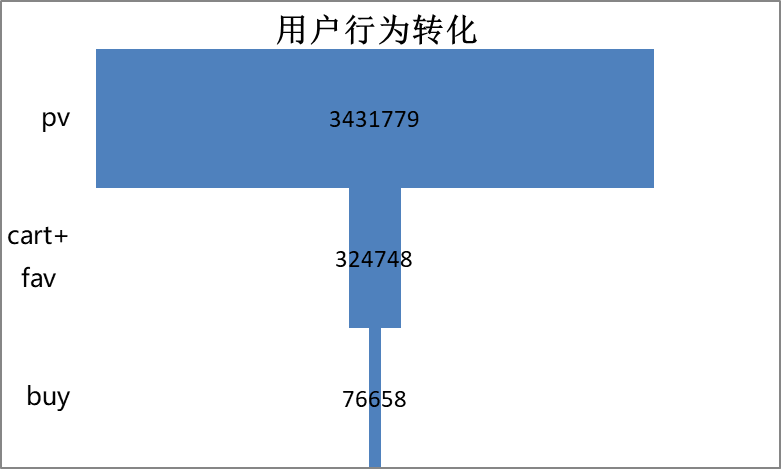


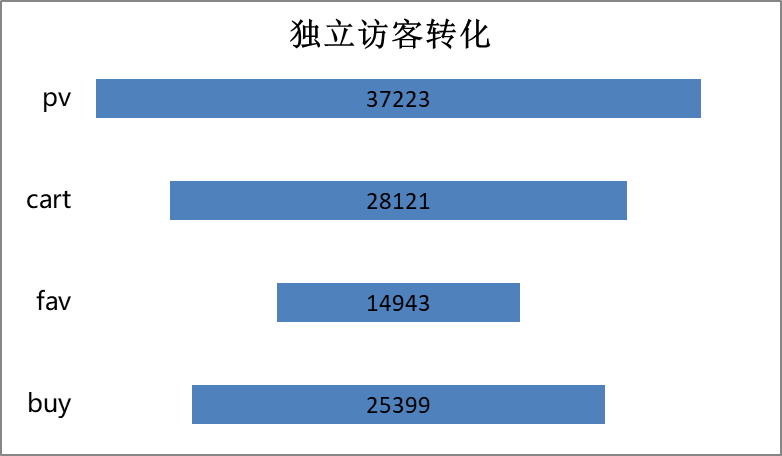
分析：支付转化率在27日-30日之间较高，30日后转化率迅速下跌。

1. 转化漏斗



用户行为转化中，将收藏和加入购物车的行为整合为第二阶段：



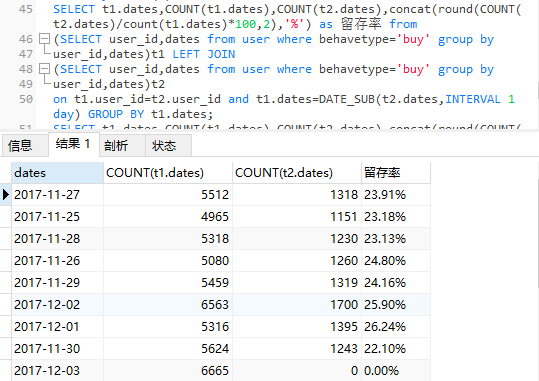


独立访客转化率达到了68%，但从用户行为来看，浏览到购买的转化率只有2.23%，若要了解具体环节的转化率，需要对整个流程接触到的版面进行细化的数据分析。

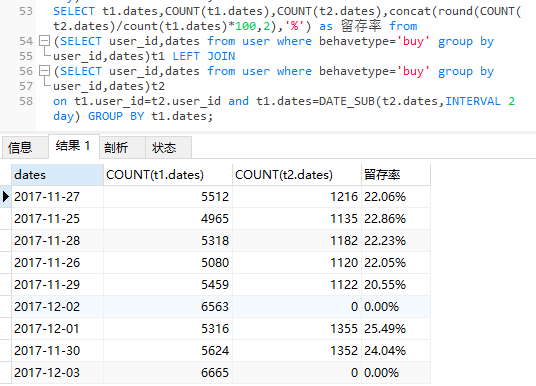
2、用户分析

（1）留存

次日留存回购人数：

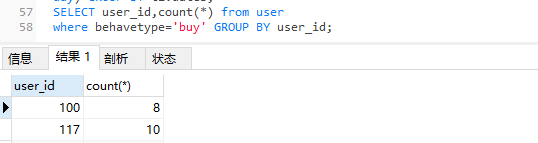


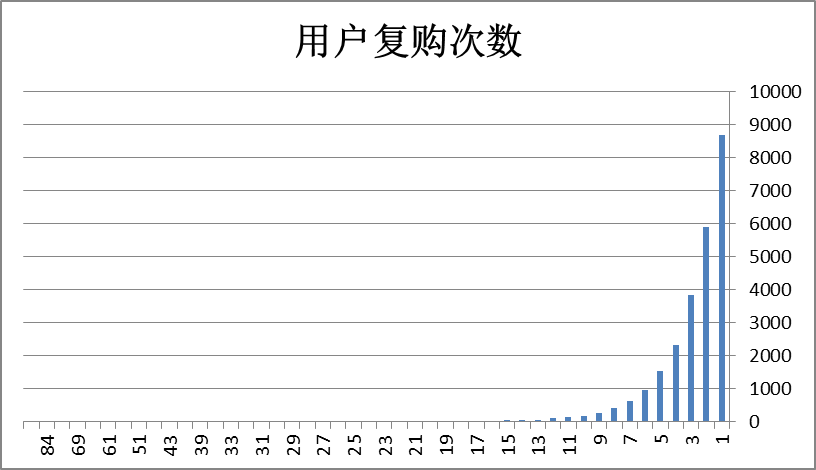
当天购买后，3天后再次购买的人数：



整体上看，次日留存率在22%-26%之间，3日留存率在22%-25%之间，说明在数据集时间范围内回购率还是很高的。

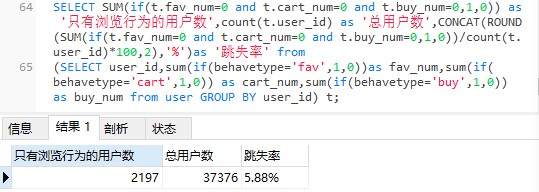
（2）复购：用户复购次数





用户复购次数集中在1-5次，呈长尾分布，该阶段用户消费欲望不大。

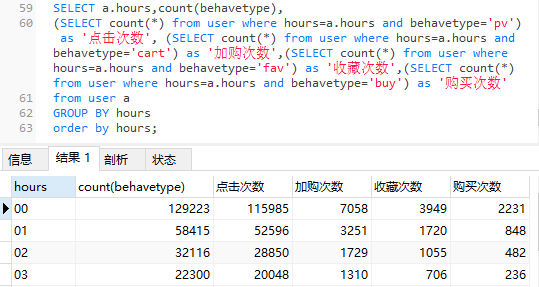
（3）跳失率：只有浏览行为的用户数/总用户数

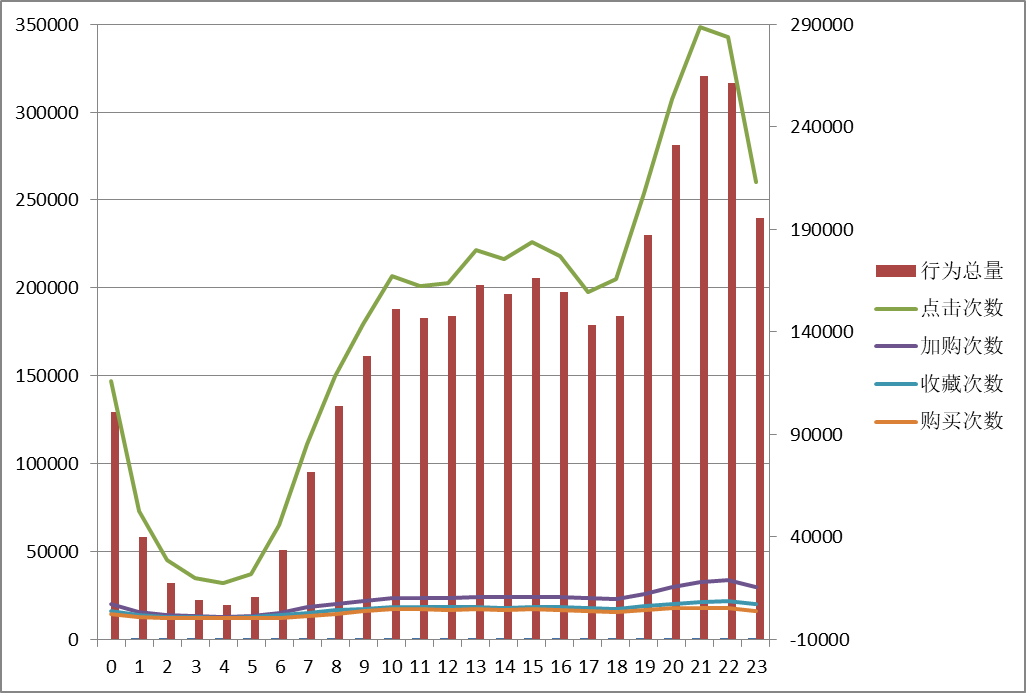


跳失率仅有5.88%，说明页面对用户的吸引力较强。

（4）用户消费习惯分析

一天中用户活跃时段分布：





可以看出：

1）每日0点到4点，用户活跃度快速降低，降到一天中的最低值，5点到10点用户活跃度逐渐上升；

2）下午整体时段的活跃度较平稳，4点到6点有所回落，该时间段处在下班、下课、吃晚饭的状态，符合用户生活习惯；

3）用户整体在晚上活跃度高，在9点-10点达到一天中的最高峰，活跃度是上午的2倍左右。

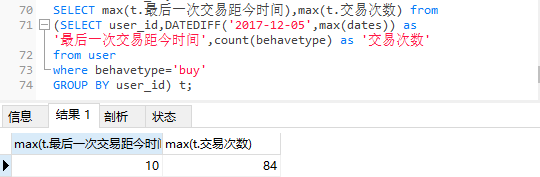
（5）用户价值分析

借鉴RFM模型，暂时不考虑M维度（由于数据集中没有给出具体金额），对R（最近一次购买时间）F（购买频率）进行分析，完成用户分层。

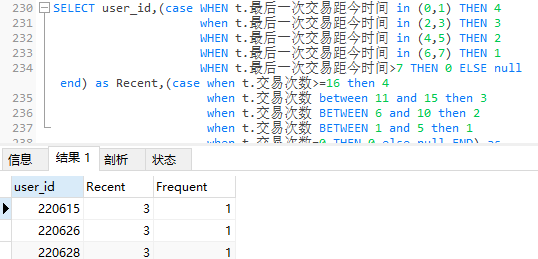
1. 最近一次消费时间和消费频率



最后一次交易距今时间和交易次数的最大值：



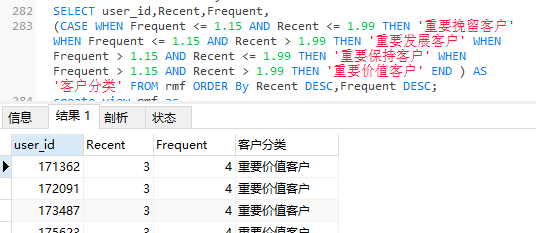
1. 对用户进行评分



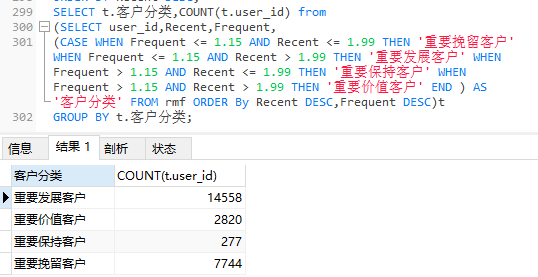
导出计算得R的平均值为1.99，F的平均值为1.15。用均值来划分4个客户层次：

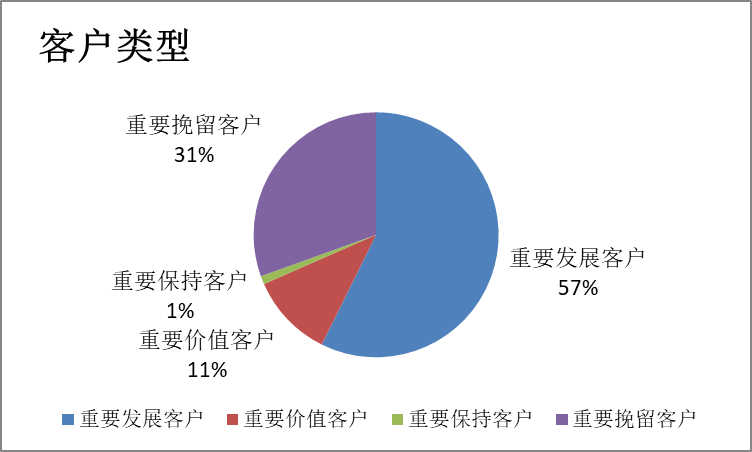
|  |  |  |
| --- | --- | --- |
| F（3-4） | 重要保持客户 | 重要价值客户 |
| F（1-2） | 重要挽留客户 | 重要发展客户 |
|  | R（1-2） | R（3-4） |

1. 客户分层绘图



各类客户数量：

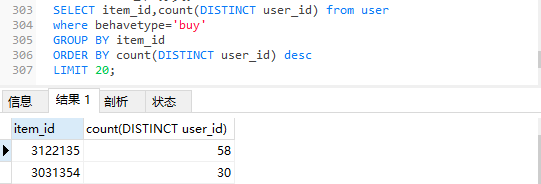


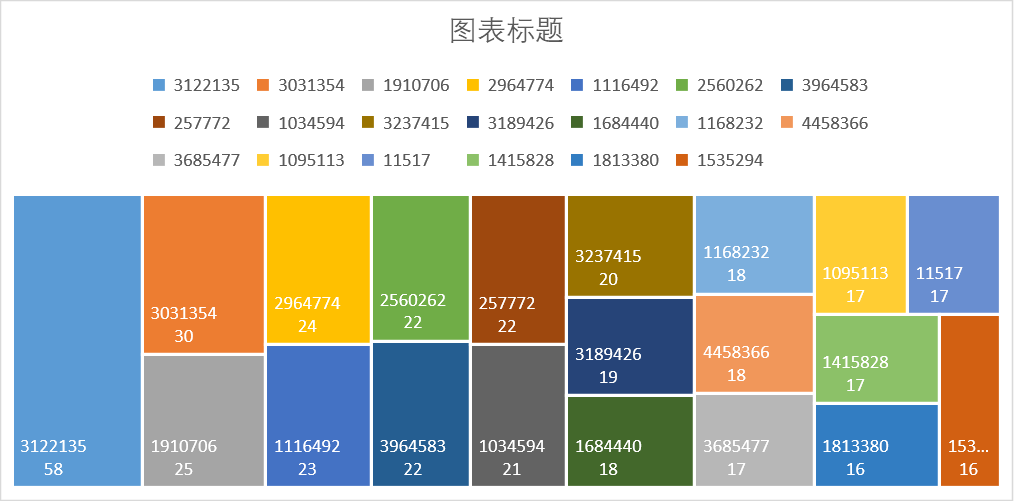


3、商品销售分析

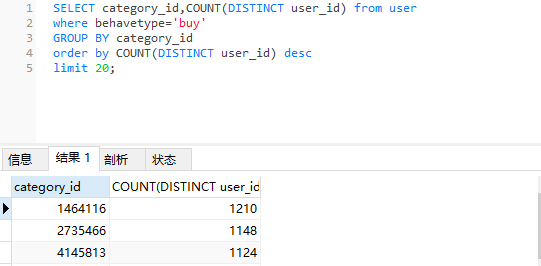
1）热销商品、热销类别

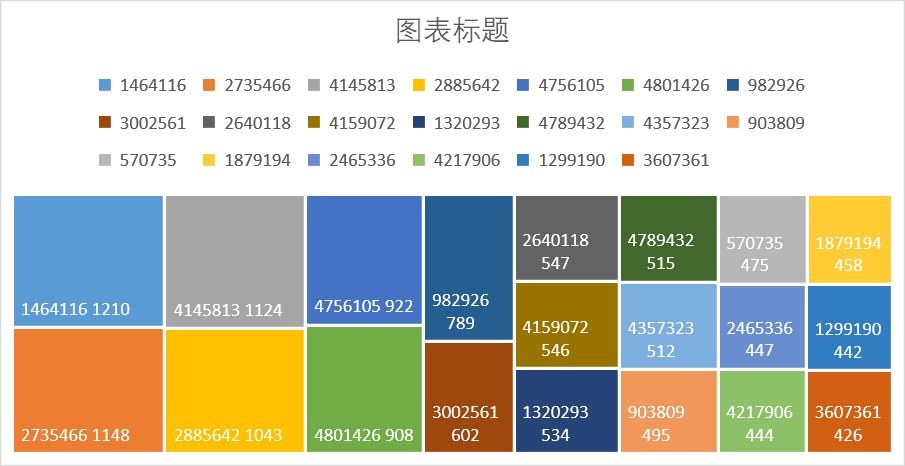
查看销量TOP20的商品：其中销量前三的商品为3122135、3031354、1910706



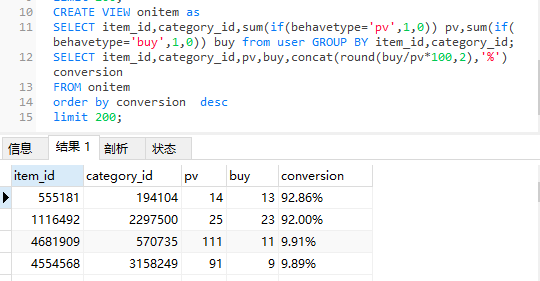


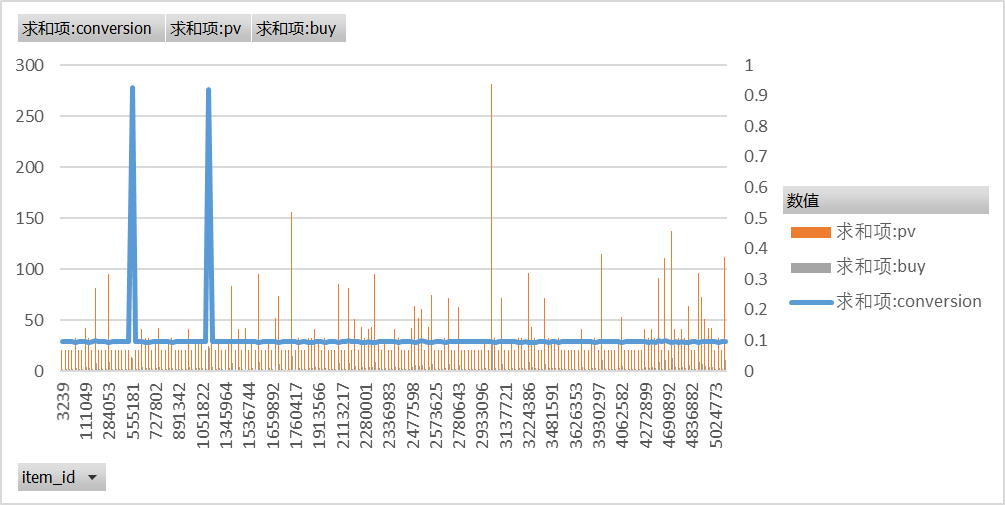
从类别上看，1464116标签号类别销量最好，共有1210个用户购买该类别商品：





2）商品购买转化率





商品编号为555181和1116492的商品购买转化率达到了92%，猜测可能是由于这两种商品在此期间进行了一系列促销活动导致的。对于其他商品浏览量高，而转化率非常低的，建议商家对商品购买流程使用转化漏斗进行分析并进行改善。