京东用户行为分析报告

姓名：黄泽森

邮箱：json.wong.work@gmail.com

目 录

[1. **前言** 1](#_Toc59807552)

[1.1 分析背景 1](#_Toc59807553)

[1.2 数据来源 1](#_Toc59807554)

[1.3 分析工具 1](#_Toc59807555)

[1.4 数据介绍 1](#_Toc59807556)

[2. **分析目的及框架** 2](#_Toc59807557)

[2.1 提出问题 2](#_Toc59807558)

[2.2 分析框架 2](#_Toc59807559)

[3. **结论汇总** 3](#_Toc59807560)

[3.1 整体分析结论 3](#_Toc59807561)

[3.2 时间维度分析结论 3](#_Toc59807562)

[3.3 用户分析结论 4](#_Toc59807563)

[3.4 商品分析结论 5](#_Toc59807564)

[4. **整体情况分析** 6](#_Toc59807565)

[4.1 整体情况分析框架及相应指标 6](#_Toc59807566)

[4.2 流量情况 6](#_Toc59807567)

[4.3 销售情况 7](#_Toc59807568)

[4.4 小结 8](#_Toc59807569)

[5. **时间维度分析** 10](#_Toc59807570)

[5.1 时间维度分析框架及相应指标 10](#_Toc59807571)

[5.2 按日期分析 10](#_Toc59807572)

[5.3 按时段分析 12](#_Toc59807573)

[5.4 小结 14](#_Toc59807574)

[6. **用户分析** 15](#_Toc59807575)

[6.1 用户分析框架及相应指标 15](#_Toc59807576)

[6.2 整体用户 15](#_Toc59807577)

[6.3 消费用户 22](#_Toc59807578)

[6.4 小结 25](#_Toc59807579)

[7. **商品分析** 27](#_Toc59807580)

[7.1 商品分析框架及相应指标 27](#_Toc59807581)

[7.2 整体情况 27](#_Toc59807582)

[7.3 商品指标与下单数的关系 27](#_Toc59807583)

[7.4 热门复购商品 30](#_Toc59807584)

[8. **附录：各分析模块SQL代码** 32](#_Toc59807585)

[8.1 整体情况 32](#_Toc59807586)

[8.2 时间维度分析 37](#_Toc59807587)

[8.3 用户分析 40](#_Toc59807588)

[8.4 商品分析 48](#_Toc59807589)

1. 前言
   1. 分析背景

本次分析根据京东2018年4月上半月（1-15日）用户行为数据进行指标分析，了解该段时间运营状况，发掘用户行为规律，找出问题并提出建议。

* 1. 数据来源

本次分析的数据来源于[JDATA竞赛](https://jdata.jd.com/html/detail.html?id=8)的用户行为数据表，并截取其中日期为2018年4月1日-2018年4月15日的用户行为数据作为此次分析的数据。

* 1. 分析工具

本次分析将数据集导入 MySQL 8.0 并使用Navicat Premium 15客户端进行数据查询，查询结果的可视化图表则使用FineBI绘制。

* 1. 数据介绍

导入MySQL的数据表字段描述如下：

|  |  |  |
| --- | --- | --- |
| 字段 | 含义 | 字段类型 |
| user\_id | 用户唯一标识 | int |
| sku\_id | 商品唯一标识 | int |
| type | 行为类型：1.浏览；2.下单；3.关注；4.评论；5.加购物车 | tinyint |
| day | 行为日期 | tinyint |
| hour | 行为时段 | tinyint |

表1.1 2018年4月上半月用户行为数据表（April\_action）字段描述

1. 分析目的及框架
   1. 提出问题
2. 从整体角度：流量指标和销售过程中各项指标及转化率是多少？
3. 从时间角度：按日期和按时段分析访问用户、浏览数和下单数的变化？
4. 从用户角度：用户的各项指标情况？各行为路径转化率? 用户数量变化及留存?消费用户的指标情况？消费次数分布及消费次数搞得客户？
5. 从商品角度：商品的各项指标情况？商品销量与指标的关系？热销商品是哪些?
   1. 分析框架

根据分析目的，本次分析的整体框架如下：

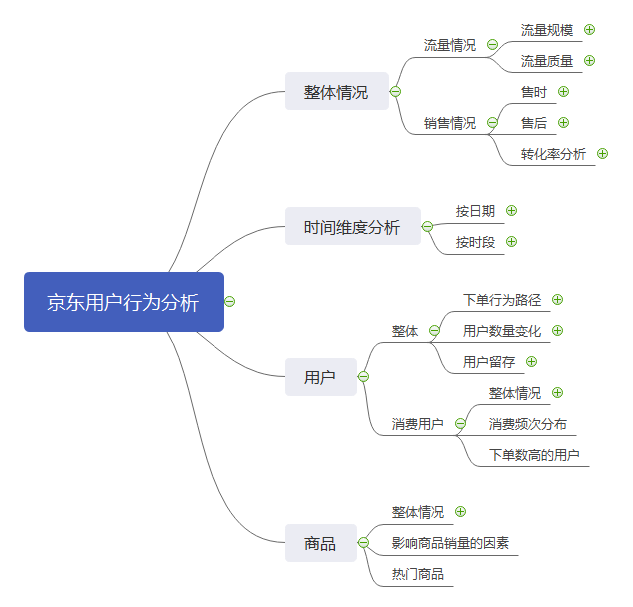


图2.1 用户行为分析框架

1. 结论汇总
   1. 整体分析结论

结论1：跳失率较高，为22.93%。

建议1：完善UI布局和操作体验，提高页面设计和内容吸引力，给与用户更好的第一印象。

结论2：浏览转化率低，仅12%。

建议2：

（1）优化商品页面的界面排布和信息介绍；

（2）优化推荐系统，提高商品和用户需求的匹配度；

（3）在用户浏览过程中创建场景，诱导产生进一步行为。如：提示用户加购/关注可即时获得价格变动、活动信息。。

* 1. 时间维度分析结论

结论1：每日10时左右和21、22时是每天活跃用户最多的时间。

建议1：

1. 在该时段进行活动和商品推送，如：新品上线、商品限时折扣；
2. 在该时段向用户发放优惠券，如：如促销或库存类商品优惠券。

结论2：每日20~22时用户平均浏览数增加，浏览量当日最高，但下单转化率较低。

建议2：

1. 提高推荐系统商品推送精准度；
2. 在该时段进行活动和商品推送，如：新品上线、商品限时折扣；
3. 在该时段向用户发放优惠券，如：如促销或库存类商品优惠券。
   1. 用户分析结论

结论 1：关注功能使用频率低且转化率低。

建议1：

1. 丰富关注功能所能提供的信息与功能；
2. 优化关注到下单之间的操作流程。

结论2：日新增用户逐日减少。

建议2：调整拉新措施，根据当前活跃用户的信息，确定目标拉新群体的普遍特征，寻找合适的渠道进行推广引流。

结论3：老用户流失加快。

建议3：

1. 福利召回。如：优惠券、抽奖礼品、等活动；
2. 社交召回。如：商品评价邀请、好友优惠助力邀请：
3. 新活动、新功能召回。如：新活动推送、新玩法。

结论4：用户留存低

建议4：

1. 培养用户日常习惯。如：日常签到奖励、连续登录奖励；
2. 鼓励用户行为。如：用户等级；
3. 在特定时间（节日/生日）推送相关内容；
4. 用户关注/购物车信息变动时及时通知。

结论5：登陆天数的越长，平均下单数越大。

结论6：用户复购率较低。

建议6：

1. 提高售时售后的服务质量，提升用户体验；
2. 推荐系统精准商品推送；
3. 根据用户关注/购买内容，及时推送相关活动；
4. 老用户回馈。如：优惠券、积分抽奖。
   1. 商品分析结论

结论1：商品从浏览到关注/加购的转化率比较低。

建议1：

（1） 提高推荐系统商品推送的精准度；

（2） 提高商品页的页面设计和内容吸引力。

结论2：评论数高的商品更能吸引用户，提高商品下单率。

建议2：

（1） 优化评论奖励机制，培养用户评论行为；

（2） 优化评论展现方式，使用户更容易获取评论数及关键词。

建议3：研究热门商品吸引用户复购的点（商品本身特点、价格变动），以借鉴推广到其他商品。

1. 整体情况分析
   1. 整体情况分析框架及相应指标

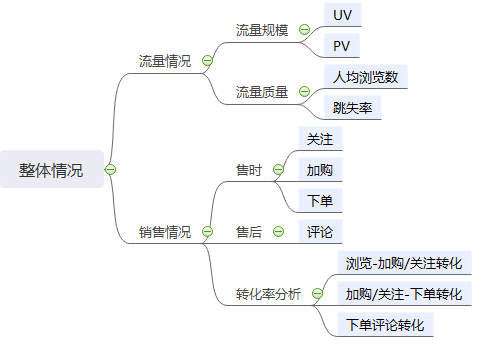


图4.1 整体情况分析框架及相应指标

* 1. 流量情况

4.1.1 流量规模



图4.2 总UV和总PV查询结果

4.1.2 流量质量



图4.3 用户平均浏览数和跳失率查询结果

可见，用户平均浏览数为8.26，而跳失率较高，为22.93% 。在不考虑误触的情况下，用户仅浏览网站一次即离开，说明用户对首页或商品页的初印象不佳，因而没有进一步操作的兴趣。因此想要降低跳失率，可以对页面设计、操作体验、内容表达等进行改善，以提升的第一印象，增加用户进一步操作的概率。

* 1. 销售情况

4.2.1 售时情况



图4.4 用户关注/加购/下单及人均情况查询结果

用户关注/加购/下单平均数均小于1，明显低于用户平均浏览数 8.26，可见用户的操作以浏览居多，而关注、加购、下单、评论的操作较少，表明用户浏览体验尚可，但进一步行为的转化率较低，需要引导用户产生更进一步的行为。

4.2.2 售后情况



图4.5 用户评论情况查询结果

4.2.3 各环节转化

由于关注行为可以发生在浏览与下单之间、加购前后，因此将加购与关注合并为一个环节。



图4.6 用户行为各环节转化情况查询结果

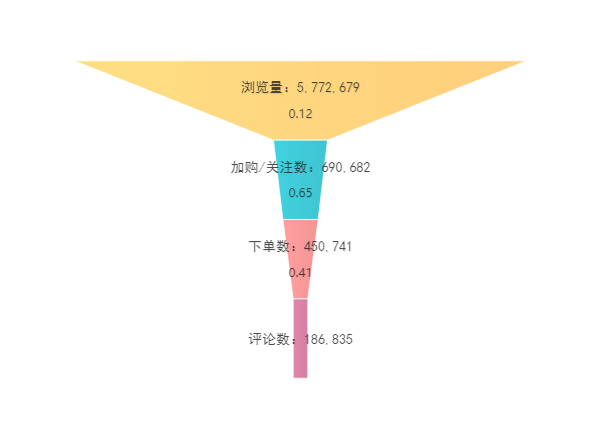


图4.7 用户行为整体指标转化漏斗

可见：

1. 浏览-加购/关注环节：仅有12%的转化率，说明用户对浏览到的的很多商品没有进一步行为的欲望，需要进一步分析是商品页面差还是商品推荐匹配度低的原因；
2. 加购/关注-下单环节：转化率为 65%，说明用户加购/关注的商品有很大概率购买；
3. 下单-评论转化率为 41%，可以进一步采取评论积分等措施提高，以形成良好的购物反馈，既能丰富商品相关信息，又能吸引更多的用户关注。
   1. 小结

结论1：跳失率较高，为22.93%。

建议1：完善UI布局和操作体验，提高页面设计和内容吸引力，给与用户更好的第一印象。

结论2：浏览转化率低，仅12%。

建议2：

（1）优化商品页面的界面排布和信息介绍；

（2）优化推荐系统，提高商品和用户需求的匹配度；

（3）在用户浏览过程中创建场景，诱导产生进一步行为。如：提示用户加购/关注可即时获得价格变动、活动信息。

1. 时间维度分析
   1. 时间维度分析框架及相应指标

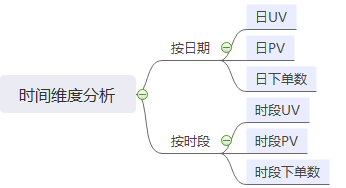


图5.1 时间维度分析框架及相应指标

* 1. 按日期分析



图5.2 每日 UV、PV、下单数查询结果

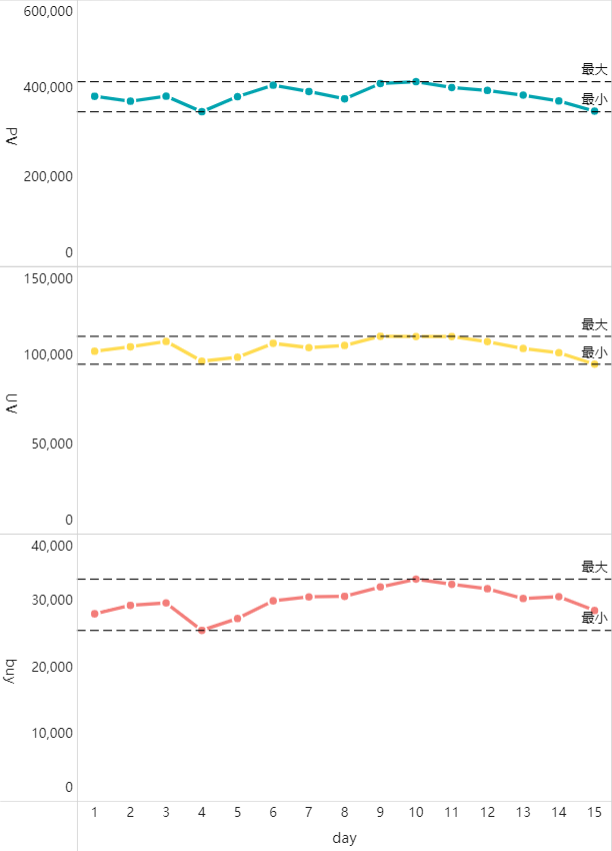


图5.3 各项指标每日变化趋势基本同步

由图表可见：

（1）4号是一个明显的低谷，当天是清明假期的前一天；

（2）8号复工至9号、10号各项指标提升并达到15天内最大值；

（3）10号过后各项指标呈逐渐降低趋势，直至15号。

注：1、8、15是周日；5、6、7号三天是是清明假期，8号复工；15号是休息日。

以上可尝试作出的假设有：

（1）4号是长假前一天，假期前夕用户出行或工作量大导致各项指标下降；

（2）每周各项指标从周一逐日提高至周三达最高点，然后逐日降低，至休息日降低至每周最低点。

但由于数据时间区间只有两周，且有一周包含了3天假期和工作日调整，故无法判断指标的周规律，而只有一个假期也无法判定假期对指标的影响，此外，由于无法得知该段时间业务方是否在某天举办了活动，故需要更长的时间周期数据支撑，进行对比分析才能验证以上推断。

* 1. 按时段分析

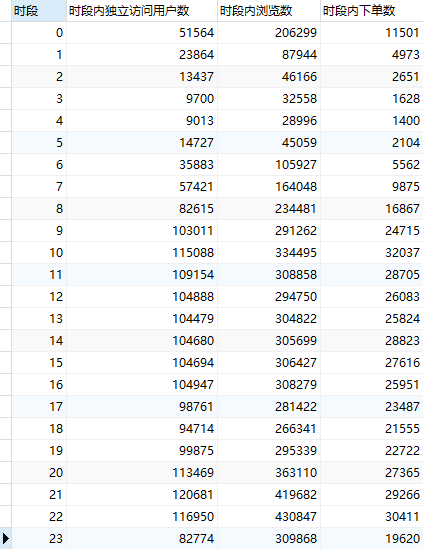


图5.4 各时段PV/UV/下单数查询结果

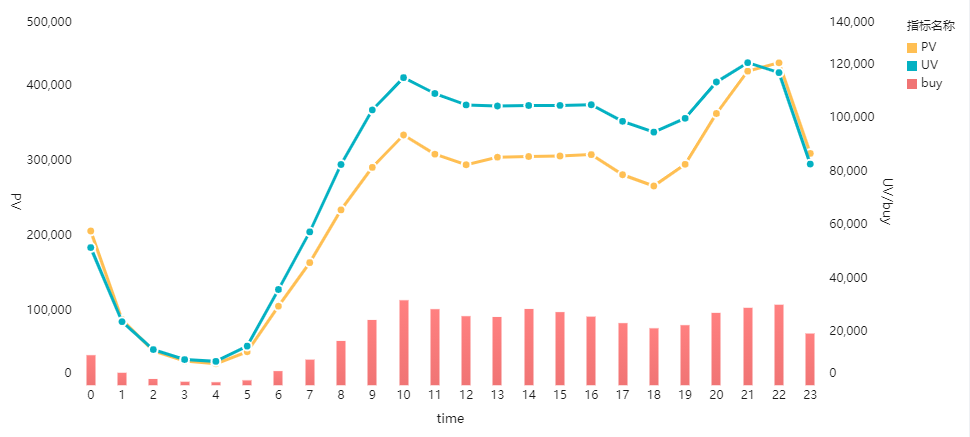


图5.5 各时段PV/UV/下单数变化

由图表可见，各时段指标的变化整体上可以分四个阶段：

1. 23时至次日4时：各项指标都在快速下降，1时过后下降速度逐渐趋缓，于4时达到当日最低点
2. 5至10时：各项指标都开始攀升，并在10时达到早高峰；
3. 11 时至16时：各项指标整体较为平稳；
4. 17时至18时：各项指标有小幅下降，并在18时降至早高峰和晚高峰之间的最低点；
5. 19时至22时：各项指标逐渐提升，并在21、22时达到晚高峰。其中，PV的在20~22时明显高于其他时段（包括早高峰），而UV和下单数的晚高峰和早高峰水平接近。

以上可得：

（1）除去深夜时间，10时和21、22时是每天用户最多也最活跃的时间，相对而言，17时至18时是相对不活跃的时间；

（2）20~22是PV明显高于其它时间，但UV并没有明显提高，说明该时段内用户平均浏览数增加；

（3）20~22是PV明显高于其它时间，但下单数并没有明显提高，说明该时段的高浏览量没能有效地转换成下单数。

* 1. 小结

结论1：每日10时左右和21、22时是每天活跃用户最多的时间。

建议1：

1. 在该时段进行活动和商品推送，如：新品上线、商品限时折扣；
2. 在该时段向用户发放优惠券，如：如促销或库存类商品优惠券。

结论2：每日20~22时用户平均浏览数增加，浏览量当日最高，但下单转化率较低。

建议2：

1. 提高推荐系统商品推送精准度；
2. 在该时段进行活动和商品推送，如：新品上线、商品限时折扣；
3. 在该时段向用户发放优惠券，如：如促销或库存类商品优惠券。
4. 用户分析
   1. 用户分析框架及相应指标

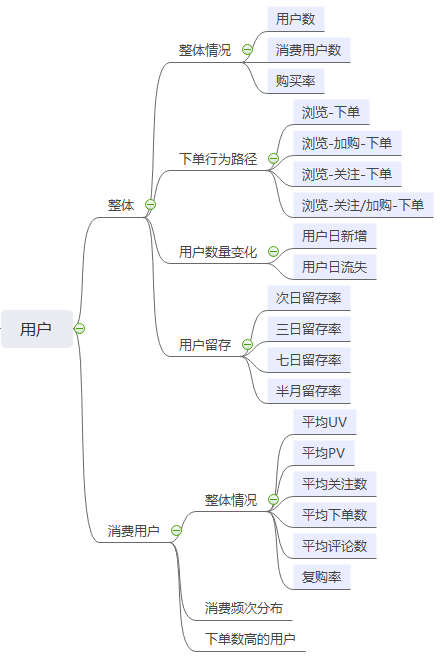


图6.1 用户分析框架及相应指标

* 1. 整体用户
     1. 整体情况



图6.2 整体用户的情况

* + 1. 下单行为路径



图6.3 路径：浏览-下单



图6.4路径：浏览-加购-下单



图6.5 路径：浏览-关注-下单



图6.6 路径：浏览-加购且关注-下单

可见：在购买用户的行为路径中，转化率最高的是“浏览-下单”，转化率为0.296；其次为路径：浏览-加购-下单，专化率为：0.1870；再者是“浏览-加购且关注-下单”，转化率为0.0296；最后是“浏览-关注-下单”，转化率为0.0213。

结论：

（1）比起关注后再下单，更多的用户选择加购后再下单，可能是因为购物车页面比起关注页面所能获得的信息更丰富、操作更便捷；

（2）下单前有关注行为的用户只有5%，可能是因为关注后再进行下一步操作的流程较为复杂。

* + 1. 用户数量变化

以4月1日至4月15日为时间区间，4月1日至4月15日有过行为记录的用户为对象，将每日第一次出现用户定义为当日新增（除1号），将每日最后一次支付用户定义为次日流失用户（除15号），查找对象在时间区间内的每日新增/流失情况如下：

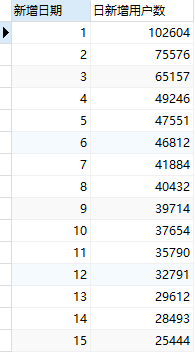
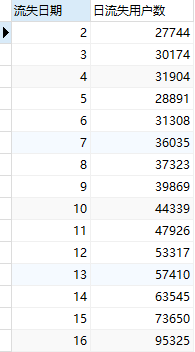
 

图6.7 用户日新增/流失数查询结果

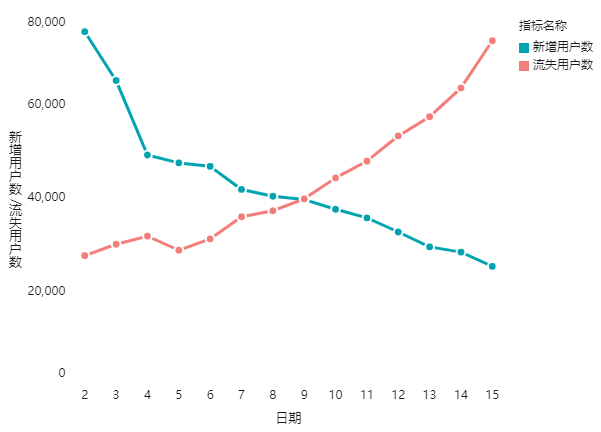


图6.8 日新增用户和流失用户数目

由图可见：

1. 每日新增用户在2号到4号期间快速减少，然后速度减缓稳定减少直至15号；
2. 每日流失用户除了在4号至7号有一个小低谷（降低再升高），流失数量一直在增加，且7号过后流失速度有加快的趋势；
3. 9号过后的每日流失用户数大于每日新增用户数，且流失速度大于新增速度。

分析：

（1）2号到4号是清明节前三天，可能是为节日作购物准备，每日新增用户较平日增多，而4号过后的每日新增趋势才是正常的；

（2）部分用户在节2号到4号期间进行节日购物之后便没有再登录过京东，所以2号到4号的流失用户数较平日较高，之后恢复到正常流失水平。

结论：

1. 日新增用户逐日减少，需要调整拉新措施；
2. 老用户流失加快，可通过老用户福
3. 利等活动等进行挽留和召回。
   * 1. 用户留存



图6.9 每日的次日/三日/七日/半月留存查询结果

图6.10 每日的次日/三日/七日/半月留存

注：由于数据日期区间为1号至15号，1号为数据日期头一天不好判定新增用户，15号不存在相关指标。

由图表可知：

1. 2号的次日留存率明显较高，推测是由于2、3号是清明假期前的一个下单小高峰，因此；
2. 从 1号到8号的各项留存都处于小幅波动的状态，较为稳定。但由于9号过后日流失用户数大于新增用户数，且流失速度大于新增速度，意味者有旧用户的减少且越来越多，于是次日留存、三日留存、七日留存分别在在9号、8号、4号过后开始有所降低。
3. 正常情况下，次日留存率在18%～19%的水平，三日留存在14～15%的水平，七日留存在12%左右的水平，比较低，但小部分用户粘性高。

结论：对于购物类产品，大多数用户只在有购物需求时才会使用，难以像社交、游戏产品产生持续性行为，因而难有较高的留存率。若要增加用户粘性，可以通过设计类似连续登录奖励机制、需要用户间互动的活动，来增加用户的留存率。

* + 1. 登陆天数与下单数的关系

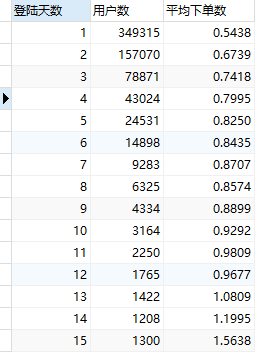


图6.11 用户登陆天数对应的用户数与平均下单数查询结果

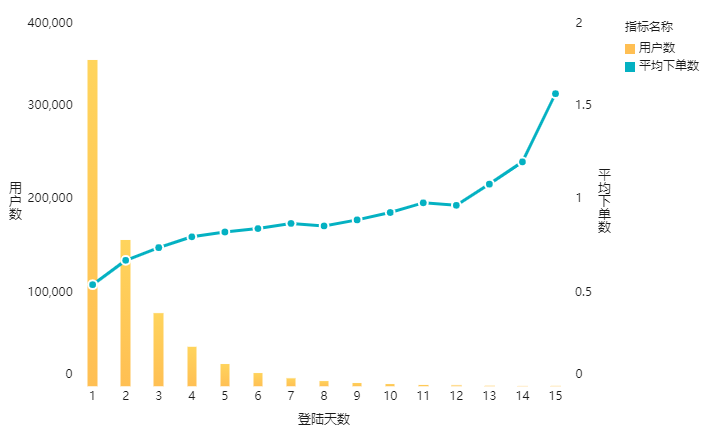


图6.12 用户登录天数及对应用户数、平均下单数

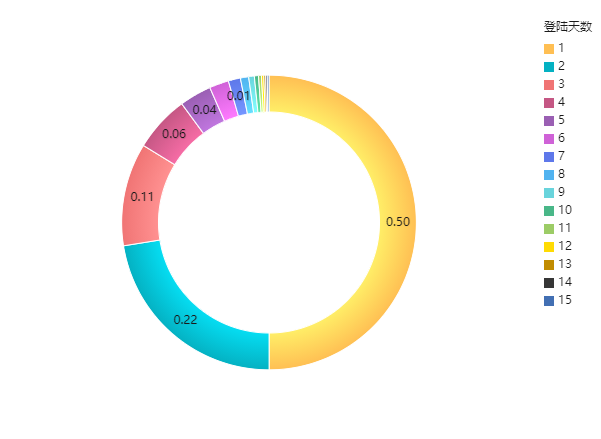


图6.13 用户的登陆天数分布

可见：

（1）登陆天数的越长，平均下单数越大；

（2）50%的用户登陆天数为1，90%的用户登陆天数在4天以内，只有10%的用户登陆天数大于4天。

结论：需要针对登陆天数少的用户采取措施进行优化。用户登陆天数少的原因可能有多种：除去用户本身没有需求，还有用户使用体验不佳、对内容不感冒等。所以有优化主要分为两部分：一方面优化界面和操作，提高用户的好感；另一方面通过商品精准推荐、定期活动推送等方式召回登陆天数少的用户，以增强用户粘性。

* 1. 消费用户
     1. 整体情况



图6.14 消费用户复购率及各项指标平均数

可见：

1. 消费用户人均浏览数高于人均浏览数，说明消费用户的浏览次数普遍高于不消费的用户；
2. 消费用户依旧存在浏览到进一步操作的转化率较低的现象；
3. 人均加购数 + 人均关注数 ≈ 人均下单数，说明消费用户只要加购或关注就很可能会下单；
4. 消费用户复购率较低，只有 13%。
   * 1. 消费频次分布

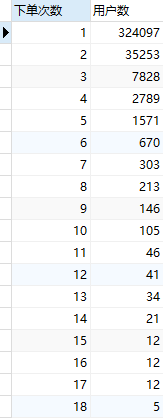
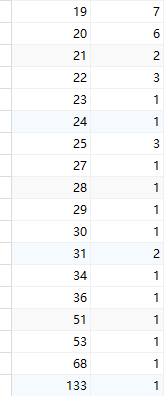
 

图6.15 用户下单次数分布查询结果

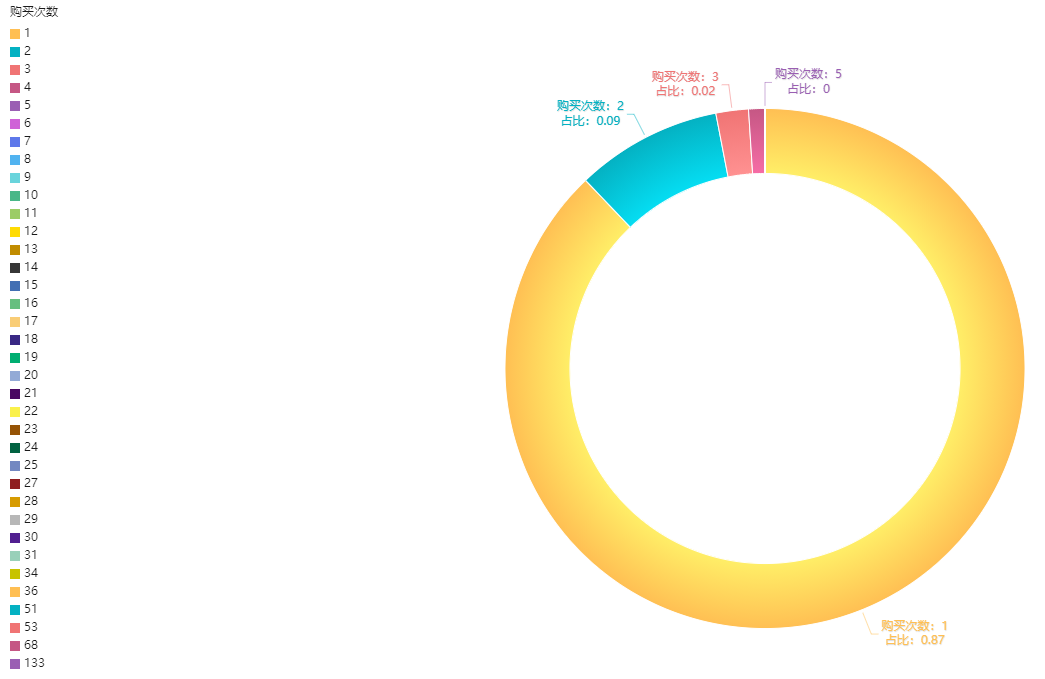
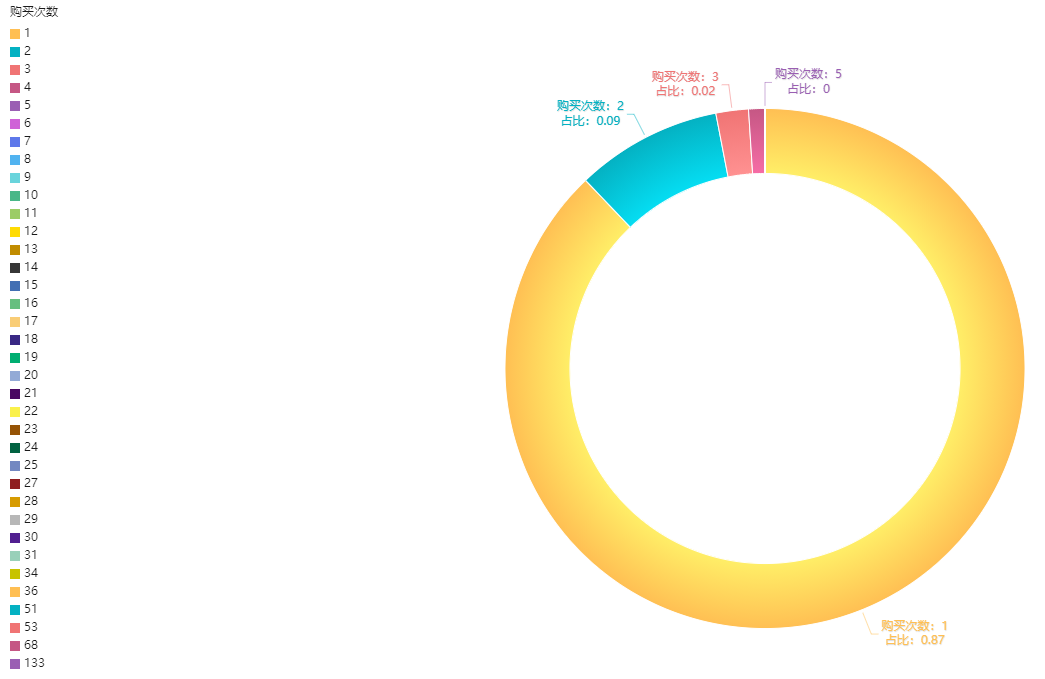


图6.16 用户下单次数占比

可见，15天内，只购买1次的用户占比87%，购买2次的用户占比9%，购买3次的用户占比2%，购买4次的用户占比1%，购买5次及以上的用户占比不到1%，而最高购买次数达133。

* + 1. 下单数前十的用户

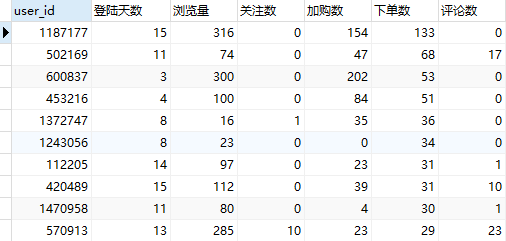


图6.17 下单数前十的用户查询结果

可以发现下单数高的用户有以下特征：

1. 累计登录天数和浏览数至少有一项比较高；
2. 更倾向于加购而不是关注；
3. 大多数不评论商品，可能是因为待评商品过多。
   1. 小结

结论 1：关注功能使用频率低且转化率低。

建议1：

1. 丰富关注功能所能提供的信息与功能；
2. 优化关注到下单之间的操作流程。

结论2：日新增用户逐日减少。

建议2：调整拉新措施，根据当前活跃用户的信息，确定目标拉新群体的普遍特征，寻找合适的渠道进行推广引流。

结论3：老用户流失加快。

建议3：

1. 福利召回。如：优惠券、抽奖礼品、等活动；
2. 社交召回。如：商品评价邀请、好友优惠助力邀请：
3. 新活动、新功能召回。如：新活动推送、新玩法。

结论4：用户留存低

建议4：

1. 培养用户日常习惯。如：日常签到奖励、连续登录奖励；
2. 鼓励用户行为。如：用户等级；
3. 在特定时间（节日/生日）推送相关内容；
4. 用户关注/购物车信息变动时及时通知。

结论5：登陆天数的越长，平均下单数越大。

结论6：用户复购率较低。

建议6：

1. 提高售时售后的服务质量，提升用户体验；
2. 推荐系统精准商品推送；
3. 根据用户关注/购买内容，及时推送相关活动；
4. 老用户回馈。如：优惠券、积分抽奖。

1. 商品分析
   1. 商品分析框架及相应指标

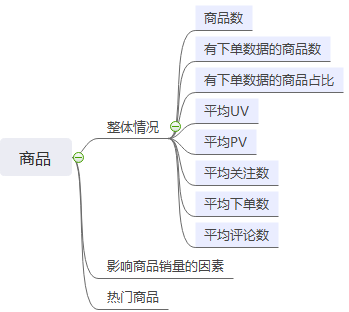


图7.1 商品分析框架及相应指标

* 1. 整体情况



图7.2 商品整体情况查询结果

可见商品从浏览到关注/加购的转化率比较低，表明向用户推荐的商品与用户实际需求的匹配度较低，需要提高商品推送的精准度。

* 1. 商品指标与下单数的关系

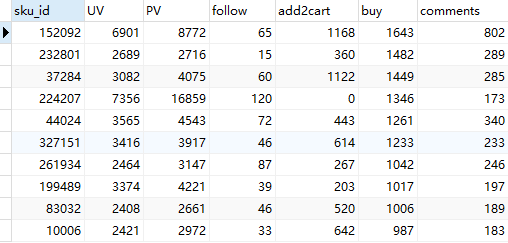


图7.3 下单数前十商品查询结果

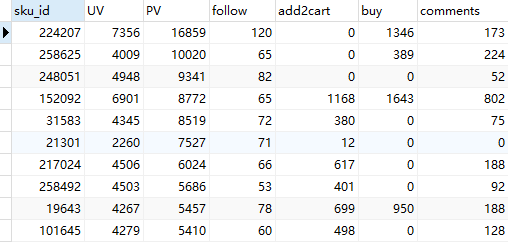


图7.4 浏览数前十商品查询结果

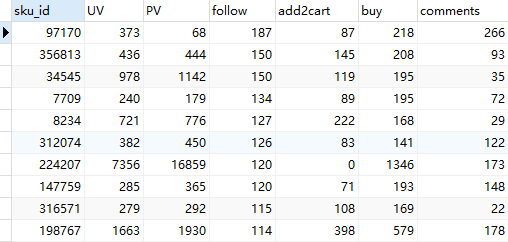


图7.5 关注数前十商品查询结果

初步发现：

1. 仅有 2 项商品同时在下单数前十和浏览数前十；
2. 仅有1项商品同时在在下单数前十和关注数前十；
3. 浏览数前十的商品中，有6项下单数为0，其评论数明显低于其他有下单数的商品（大多<100）。

注：由于加购数据的缺失，此处不作关系探索。

分别进行分析：

1. 流量高下单少，说明用户可能对能够高频浏览到的商品不感兴趣。

假设：推荐的商品与用户的需求不匹配，导致高浏览量商品没有转换成相应的下单数。

验证：取前下单数和浏览数前100的商品进行连接，查看同时在榜的商品。



发现仅有 28 件商品同时在榜，这说明浏览量高的商品到下单的转化率较低，反映了推送的商品与用户需求的匹配度并不高。

1. 关注高的商品下单数不高，说明用户关注商品的目的不一定是为了下单。

假设：高关注商品不一定是热销商品。

验证：取下单数前100商品中关注数前100的商品数。



发现仅有 15 件商品同时在榜，可见关注数不是衡量商品是否热销的指标，用户会基于除了买之外的目的对商品进行关注，假设成立。

1. 商品浏览量高，却没能明显转化成销量，而这部分商品评论数都比较低。

假设：高浏览量的商品中，高评论数的商品下单数高于低评论数商品。

下面进行验证，查看浏览量前100中，评论数高于浏览量前100商品平均数的商品下单数占比：



可见，评论数高于平均的商品下单数占比超过90%，故假设成立。

假设：销量高的商品往往伴随着高评论数。取前下单数和评论数前100的商品进行连接：



发现共有 48 件商品同时在榜。说明：下单数与评论数正相关，更多的用户评论能够更吸引浏览的消费者兴趣，增强了商品的信息度和可信度，增加下单概率，从而提高销量，形成正向循环。

综上，建议采取措施如下：

1. 优化推荐系统，提高推送的精准率，以提高浏览量商品的下单转化率；
2. 优化评论奖励机制，诱导下单用户评论，丰富商品的表面和数据相关信息，以吸引更多用户，提高商品下单率。
   1. 热门复购商品

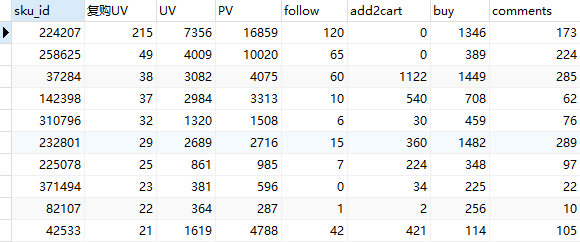


图7.6 复购用户数前十商品查询结果

针对复购用户多的商品，可以具体研究其吸引用户复购的点（商品本身特点、价格变动）以借鉴推广到其他商品。

* 1. 小结

结论1：商品从浏览到关注/加购的转化率比较低。

建议1：

1. 提高推荐系统商品推送的精准度；
2. 提高商品页的页面设计和内容吸引力。

结论2：评论数高的商品更能吸引用户，提高商品下单率。

建议2：

1. 优化评论奖励机制，培养用户评论行为；
2. 优化评论展现方式，使用户更容易获取评论数及关键词。

建议3：研究热门商品吸引用户复购的点（商品本身特点、价格变动），以借鉴推广到其他商品。

1. 附录：各分析模块SQL代码
   1. 整体情况

DROP TABLE IF EXISTS t\_user\_behaviour\_tmp;

DROP TABLE IF EXISTS t\_bounce\_tmp;

-- 统计用户行为的临时表

CREATE TEMPORARY TABLE t\_user\_behaviour\_tmp

SELECT

COUNT(DISTINCT user\_id) AS UV,

SUM(IF(type=1, 1, 0)) AS browse,

SUM(IF(type=2, 1, 0)) AS buy,

SUM(IF(type=3, 1, 0)) AS follow,

SUM(IF(type=4, 1, 0)) AS review,

SUM(IF(type=5, 1, 0)) AS add2cart

FROM

april\_actions;

-- 统计跳失用户的临时表

CREATE TEMPORARY TABLE t\_bounce\_tmp

SELECT

COUNT(\*) AS bounce\_n

FROM

(

SELECT

user\_id

FROM

april\_actions

GROUP BY

user\_id

HAVING

COUNT(\*) = 1

) AS t\_bounce;

-- 流量规模

SELECT

t\_user\_behaviour\_tmp.UV AS '独立访客数(UV)',

t\_user\_behaviour\_tmp.browse AS '浏览量(PV)'

FROM

t\_user\_behaviour\_tmp;

-- 流量质量

SELECT

ROUND(t\_user\_behaviour\_tmp.browse / t\_user\_behaviour\_tmp.UV, 2) AS '用户平均浏览数',

ROUND(t\_bounce\_tmp.bounce\_n / t\_user\_behaviour\_tmp.UV \* 100, 2) AS '跳失率(%)'

FROM

t\_user\_behaviour\_tmp,

t\_bounce\_tmp;

-- 售时

SELECT

t\_user\_behaviour\_tmp.follow AS '关注数',

ROUND(t\_user\_behaviour\_tmp.follow / t\_user\_behaviour\_tmp.UV, 2) AS '用户平均关注数',

t\_user\_behaviour\_tmp.add2cart AS '加购数',

ROUND(t\_user\_behaviour\_tmp.add2cart / t\_user\_behaviour\_tmp.UV, 2) AS '用户平均加购数',

t\_user\_behaviour\_tmp.buy AS '下单数',

ROUND(t\_user\_behaviour\_tmp.buy / t\_user\_behaviour\_tmp.UV, 2) AS '用户平均下单数'

FROM

t\_user\_behaviour\_tmp;

-- 售后

SELECT

t\_user\_behaviour\_tmp.review AS '评论数',

ROUND(t\_user\_behaviour\_tmp.review / t\_user\_behaviour\_tmp.UV, 2) AS '用户平均评论数'

FROM

t\_user\_behaviour\_tmp;

-- 整体行为转化漏斗

SELECT

t\_user\_behaviour\_tmp.browse AS '浏览量',

ROUND((t\_user\_behaviour\_tmp.follow + t\_user\_behaviour\_tmp.add2cart) / t\_user\_behaviour\_tmp.browse \* 100, 2) AS '浏览-加购/关注转化率(%)',

t\_user\_behaviour\_tmp.follow + t\_user\_behaviour\_tmp.add2cart AS '加购/关注数',

ROUND(t\_user\_behaviour\_tmp.buy / (t\_user\_behaviour\_tmp.follow + t\_user\_behaviour\_tmp.add2cart) \* 100, 2) AS '加购/关注-下单转化率(%)',

t\_user\_behaviour\_tmp.buy AS '下单数',

ROUND(t\_user\_behaviour\_tmp.review / t\_user\_behaviour\_tmp.buy \* 100, 2) AS '下单-评论转化率(%)',

t\_user\_behaviour\_tmp.review AS '评论数'

FROM

t\_user\_behaviour\_tmp;

-- -- 汇总整体指标

-- SELECT

-- t\_user\_behaviour\_tmp.UV AS '独立访客数(UV)',

-- t\_user\_behaviour\_tmp.browse AS '浏览量(PV)',

-- ROUND(t\_user\_behaviour\_tmp.browse / t\_user\_behaviour\_tmp.UV, 2) AS '用户平均浏览数',

-- t\_user\_behaviour\_tmp.buy AS '下单数',

-- ROUND(t\_user\_behaviour\_tmp.buy / t\_user\_behaviour\_tmp.UV, 2) AS '用户平均下单数',

-- t\_user\_behaviour\_tmp.follow AS '关注数',

-- ROUND(t\_user\_behaviour\_tmp.follow / t\_user\_behaviour\_tmp.UV, 2) AS '用户平均关注数',

-- t\_user\_behaviour\_tmp.review AS '评论数',

-- ROUND(t\_user\_behaviour\_tmp.review / t\_user\_behaviour\_tmp.UV, 2) AS '用户平均评论数',

-- t\_user\_behaviour\_tmp.add2cart AS '加购数',

-- ROUND(t\_user\_behaviour\_tmp.add2cart / t\_user\_behaviour\_tmp.UV, 2) AS '用户平均加购数',

-- ROUND(t\_bounce\_tmp.bounce\_n / t\_user\_behaviour\_tmp.UV \* 100, 2) AS '跳失率(%)'

-- FROM

-- t\_user\_behaviour\_tmp,

-- t\_bounce\_tmp;

--

-- -- 整体行为转化漏斗

-- SELECT

-- t\_user\_behaviour\_tmp.browse AS '浏览量',

-- ROUND((t\_user\_behaviour\_tmp.follow + t\_user\_behaviour\_tmp.add2cart) / t\_user\_behaviour\_tmp.browse \* 100, 2) AS '浏览-加购/关注转化率(%)',

-- t\_user\_behaviour\_tmp.follow + t\_user\_behaviour\_tmp.add2cart AS '加购/关注数',

-- ROUND(t\_user\_behaviour\_tmp.buy / (t\_user\_behaviour\_tmp.follow + t\_user\_behaviour\_tmp.add2cart) \* 100, 2) AS '加购/关注-下单转化率(%)',

-- t\_user\_behaviour\_tmp.buy AS '下单数',

-- ROUND(t\_user\_behaviour\_tmp.review / t\_user\_behaviour\_tmp.buy \* 100, 2) AS '下单-评论转化率(%)',

-- t\_user\_behaviour\_tmp.review AS '评论数'

-- FROM

-- t\_user\_behaviour\_tmp;

* 1. 时间维度分析

DROP TABLE IF EXISTS t\_day\_trend\_tmp;

DROP TABLE IF EXISTS t\_hour\_trend\_tmp;

-- 统计每天数据

CREATE TEMPORARY TABLE t\_day\_trend\_tmp

SELECT

day,

COUNT(DISTINCT user\_id) AS UV,

SUM(IF(type=1, 1, 0)) AS PV,

SUM(IF(type=2, 1, 0)) AS buy,

SUM(IF(type=3, 1, 0)) AS follow,

SUM(IF(type=4, 1, 0)) AS review,

SUM(IF(type=5, 1, 0)) AS add2cart

FROM

april\_actions

GROUP BY

day;

-- 统计各时段数据

CREATE TEMPORARY TABLE t\_hour\_trend\_tmp

SELECT

hour,

COUNT(DISTINCT user\_id) AS UV,

SUM(IF(type=1, 1, 0)) AS PV,

SUM(IF(type=2, 1, 0)) AS buy,

SUM(IF(type=3, 1, 0)) AS follow,

SUM(IF(type=4, 1, 0)) AS review,

SUM(IF(type=5, 1, 0)) AS add2cart

FROM

april\_actions

GROUP BY

hour;

-- 查看每日 UV, PV, buy

SELECT

day AS '日期',

UV AS '日独立访问用户数',

PV AS '日浏览数',

buy AS '日下单数'

FROM

t\_day\_trend\_tmp;

-- 查看各时段 UV, PV, buy

SELECT

hour AS '时段',

UV AS '时段内独立访问用户数',

PV AS '时段内浏览数',

buy AS '时段内下单数'

FROM

t\_hour\_trend\_tmp;

-- -- 查看每日统计信息

-- SELECT

-- \*

-- FROM

-- t\_day\_trend\_tmp;

--

-- -- 查看时段统计信息

-- SELECT

-- \*

-- FROM

-- t\_hour\_trend\_tmp;

* 1. 用户分析

DROP TABLE IF EXISTS t\_user\_tmp;

DROP TABLE IF EXISTS t\_consumer\_tmp;

-- 用户信息临时表

CREATE TEMPORARY TABLE t\_user\_tmp

SELECT

user\_id,

MIN(day) AS first\_day,

MAX(day) AS recent\_day,

COUNT(DISTINCT day) AS login\_days,

SUM(IF(type=1, 1, 0)) AS browse,

SUM(IF(type=3, 1, 0)) AS follow,

SUM(IF(type=5, 1, 0)) AS add2cart,

SUM(IF(type=2, 1, 0)) AS buy,

SUM(IF(type=4, 1, 0)) AS review

FROM

april\_actions

GROUP BY

user\_id;

-- 消费用户信息临时表

CREATE TEMPORARY TABLE t\_consumer\_tmp

SELECT

user\_id,

COUNT(DISTINCT day) AS login\_days,

SUM(IF(type=1, 1, 0)) AS browse,

SUM(IF(type=3, 1, 0)) AS follow,

SUM(IF(type=5, 1, 0)) AS add2cart,

SUM(IF(type=2, 1, 0)) AS buy,

SUM(IF(type=4, 1, 0)) AS review

FROM

april\_actions

GROUP BY

user\_id

HAVING

SUM(IF(type=2, 1, 0)) > 0;

-- 用户整体情况

SELECT

@total\_n AS '用户数',

COUNT(\*) AS '消费用户数',

COUNT(\*) / @total\_n AS '用户购买率'

FROM

t\_consumer\_tmp,

(SELECT @total\_n := (SELECT COUNT(\*) FROM t\_user\_tmp)) AS var ;

-- 行为路径分析。由于分析的行为路径起点为浏览，以下行为的分析基于有浏览行为的下单用户

-- 浏览-下单

SELECT

@browse\_user\_n AS '浏览用户数',

SUM(IF(buy > 0, 1, 0)) AS '下单用户数'

FROM

t\_user\_tmp

WHERE

browse > 0 AND follow = 0 AND add2cart = 0;

-- 浏览-关注/加购-下单

SELECT

@browse\_user\_n AS '浏览用户数',

COUNT(\*) AS '关注/加购用户数',

SUM(IF(buy > 0, 1, 0)) AS '下单用户数'

FROM

t\_user\_tmp

WHERE

browse > 0 AND follow > 0 AND add2cart > 0;

-- 浏览-关注-下单

SELECT

@browse\_user\_n AS '浏览用户数',

COUNT(\*) AS '关注用户数',

SUM(IF(buy > 0, 1, 0)) AS '下单用户数'

FROM

t\_user\_tmp

WHERE

browse > 0 AND follow > 0 AND add2cart = 0;

-- 浏览-加购-下单

SELECT

@browse\_user\_n AS '浏览用户数',

COUNT(\*) AS '加购用户数',

SUM(IF(buy > 0, 1, 0)) AS '下单用户数'

FROM

t\_user\_tmp

WHERE

browse > 0 AND follow = 0 AND add2cart > 0;

-- 日新增用户数

SELECT

first\_day '新增日期',

COUNT(\*) AS '日新增用户数'

FROM

t\_user\_tmp

GROUP BY

first\_day

ORDER BY

first\_day;

-- 日流失用户数：最近登陆日期为前一天的用户数。如最近一次登陆日期为 1 号，表明用户在 2 号流失。

SELECT

recent\_day + 1 AS '流失日期',

COUNT(\*) AS '日流失用户数'

FROM

t\_user\_tmp

GROUP BY

recent\_day

ORDER BY

recent\_day;

-- 留存率

WITH

t\_retention\_n AS

(

SELECT

first\_day,

nday,

COUNT(DISTINCT user\_id) AS retention\_n

FROM

(

SELECT

uit.user\_id,

uit.first\_day,

fa.day - uit.first\_day AS nday

FROM

t\_user\_tmp AS uit

LEFT JOIN

april\_actions AS fa

ON

uit.user\_id = fa.user\_id

) AS t

GROUP BY

first\_day,

nday

ORDER BY

first\_day,

nday

)

SELECT

first\_day AS '日期',

ROUND(SUM(IF(nday = 1, retention\_n, 0)) / SUM(IF(nday = 0, retention\_n, 0)) \* 100, 2) AS '次日留存(%)',

ROUND(SUM(IF(nday = 2, retention\_n, 0)) / SUM(IF(nday = 0, retention\_n, 0)) \* 100, 2) AS '三日留存(%)',

ROUND(SUM(IF(nday = 6, retention\_n, 0)) / SUM(IF(nday = 0, retention\_n, 0)) \* 100, 2) AS '七日留存(%)',

ROUND(SUM(IF(nday = 14, retention\_n, 0)) / SUM(IF(nday = 0, retention\_n, 0)) \* 100, 2) AS '半月留存(%)'

FROM

t\_retention\_n

GROUP BY

first\_day;

-- 用户登陆天数分布及平均下单数

SELECT

login\_days AS '登陆天数',

COUNT(\*) AS '用户数',

AVG(buy) AS '平均下单数'

FROM

t\_user\_tmp

GROUP BY

login\_days

ORDER BY

login\_days;

-- 消费用户整体情况

WITH

t AS (

SELECT

COUNT(\*) AS purchase\_n,

SUM(browse) AS browse,

SUM(follow) AS follow,

SUM(add2cart) AS add2cart,

SUM(buy) AS buy,

SUM(review) AS review,

SUM(IF(buy > 1, 1, 0)) AS repurchase\_n

FROM

t\_consumer\_tmp

)

SELECT

ROUND(repurchase\_n / purchase\_n \* 100, 2) AS '复购率(%)',

browse / purchase\_n AS '平均浏览量',

follow / purchase\_n AS '平均关注数',

add2cart / purchase\_n AS '平均加购数',

buy / purchase\_n AS '平均下单数',

review / purchase\_n AS '平均评论数'

FROM

t;

-- 下单次数分布

SELECT

buy AS '下单次数',

COUNT(\*) AS '用户数'

FROM

t\_consumer\_tmp

GROUP BY

buy

ORDER BY

buy;

-- 下单数前十的用户

SELECT

user\_id,

login\_days AS '登陆天数',

browse AS '浏览量',

follow AS '关注数',

add2cart AS '加购数',

buy AS '下单数',

review AS '评论数'

FROM

t\_consumer\_tmp

ORDER BY

buy DESC

LIMIT 10;

* 1. 商品分析

DROP TABLE IF EXISTS t\_sku\_tmp;

CREATE TEMPORARY TABLE t\_sku\_tmp

SELECT

sku\_id,

COUNT(DISTINCT user\_id) AS UV,

SUM(IF(type=1, 1, 0)) AS PV,

SUM(IF(type=3, 1, 0)) AS follow,

SUM(IF(type=5, 1, 0)) AS add2cart,

SUM(IF(type=2, 1, 0)) AS buy,

SUM(IF(type=4, 1, 0)) AS comments

FROM

april\_actions

GROUP BY

sku\_id;

-- 商品的整体情况

SELECT

COUNT(\*) AS '商品数',

SUM(IF(buy > 0, 1, 0)) AS '有下单数据的商品',

SUM(IF(buy > 0, 1, 0)) / COUNT(\*) \* 100 AS '有下单数据的商品占比（%）',

AVG(UV) AS '平均访问人数',

AVG(PV) AS '平均浏览量',

AVG(follow) AS '平均关注数',

AVG(add2cart) AS '平均加购数',

AVG(buy) AS '平均下的那数',

AVG(comments) AS '平均评论数'

FROM

t\_sku\_tmp;

-- 有下单数据商品的情况

SELECT

AVG(UV) AS '平均访问人数',

AVG(PV) AS '平均浏览量',

AVG(follow) AS '平均关注数',

AVG(add2cart) AS '平均加购数',

AVG(buy) AS '平均下的那数',

AVG(comments) AS '平均评论数'

FROM

t\_sku\_tmp

WHERE

buy > 0;

-- 查看下单前十的商品

SELECT

\*

FROM

t\_sku\_tmp

ORDER BY

buy DESC

LIMIT 10;

-- 浏览量前十商品

SELECT

\*

FROM

t\_sku\_tmp

ORDER BY

PV DESC

LIMIT 10;

-- 关注数前十商品

SELECT

\*

FROM

t\_sku\_tmp

ORDER BY

follow DESC

LIMIT 10;

-- 下单数前100商品中浏览数前100商品

WITH

t\_cnt AS (

SELECT

sku\_id,

SUM(IF(type=1, 1, 0)) AS PV,

SUM(IF(type=2, 1, 0)) AS buy,

SUM(IF(type=4, 1, 0)) AS comments

FROM

april\_actions

GROUP BY

sku\_id

)

SELECT

COUNT(\*) AS '下单数前100商品中浏览数前100商品'

FROM

(

SELECT

\*

FROM

t\_cnt

ORDER BY

buy DESC

LIMIT 100

) AS b\_100

JOIN

(

SELECT

\*

FROM

t\_cnt

ORDER BY

PV DESC

LIMIT 100

) AS p\_100

ON

b\_100.sku\_id = p\_100.sku\_id;

-- 下单数前100商品中评论数前100商品

WITH

t\_cnt AS (

SELECT

sku\_id,

SUM(IF(type=1, 1, 0)) AS PV,

SUM(IF(type=2, 1, 0)) AS buy,

SUM(IF(type=4, 1, 0)) AS comments

FROM

april\_actions

GROUP BY

sku\_id

)

SELECT

COUNT(\*) AS '下单数前100商品中评论数前100商品'

FROM

(

SELECT

\*

FROM

t\_cnt

ORDER BY

buy DESC

LIMIT 100

) AS b\_100

JOIN

(

SELECT

\*

FROM

t\_cnt

ORDER BY

comments DESC

LIMIT 100

) AS c\_100

ON

b\_100.sku\_id = c\_100.sku\_id;

-- 下单数前100商品中关注数前100商品

WITH

t\_cnt AS (

SELECT

sku\_id,

SUM(IF(type=1, 1, 0)) AS PV,

SUM(IF(type=2, 1, 0)) AS buy,

SUM(IF(type=3, 1, 0)) AS follow

FROM

april\_actions

GROUP BY

sku\_id

)

SELECT

COUNT(\*) AS '下单数前100商品中关注数前100的商品'

FROM

(

SELECT

\*

FROM

t\_cnt

ORDER BY

buy DESC

LIMIT 100

) AS b\_100

JOIN

(

SELECT

\*

FROM

t\_cnt

ORDER BY

follow DESC

LIMIT 100

) AS f\_100

ON

b\_100.sku\_id = f\_100.sku\_id;

-- 浏览数前100商品中，评论数高于平均评论数的商品下单数占比

WITH

p\_100 AS

(

SELECT

sku\_id,

buy,

comments

FROM

(

SELECT

sku\_id,

SUM(IF(type=1, 1, 0)) AS PV,

SUM(IF(type=2, 1, 0)) AS buy,

SUM(IF(type=4, 1, 0)) AS comments

FROM

april\_actions

GROUP BY

sku\_id

) AS cnt

ORDER BY

PV DESC

LIMIT 100

)

SELECT

SUM(IF(tst.comments > (SELECT AVG(comments) FROM p\_100), tst.buy, 0)) / SUM(tst.buy) AS '评论数高于平均的商品下单数占比'

FROM

p\_100 AS p

JOIN

t\_sku\_tmp AS tst

ON

p.sku\_id = tst.sku\_id;

-- 复购用户前十的商品及其购买次数

WITH tt

AS (

SELECT

user\_id,

sku\_id,

COUNT(\*) AS cnt

FROM

april\_actions

GROUP BY

user\_id,

sku\_id

HAVING

SUM(IF(type=2, 1, 0)) > 1

)

SELECT

tt.sku\_id,

COUNT(DISTINCT tt.user\_id) AS '复购UV',

tst.UV,

tst.PV,

tst.follow,

tst.add2cart,

tst.buy,

tst.comments

FROM

tt

JOIN

t\_sku\_tmp AS tst

ON

tt.sku\_id = tst.sku\_id

GROUP BY

tt.sku\_id

ORDER BY

`复购UV` DESC

LIMIT 10;