

1. 项目背景

本项目数据集来源于天池算法竞赛，数据包含了 2017 年 11 月 25 日-2017 年 12 月 3 日的用户的所有行为，数据集的每一行表示一条用户行为，由用户 id，商品 id，商品类目 id，行为类型（pv，fav，cart，buy），行为时间组成。旨在通过建立指标体系，分析用户行为和商品潜力，为实现精准营销提供数据支持。

2. 数据导入

本项目通过 Dbeaver 操作 MySQL 完成数据分析。首先新建一个数据库，命名为“taobao”，然后导入 csv 格式的数据。此时数据库中包含了 userbehavior 这个表，打开此表进行数据检查，发现没有字段名，需要将字段名加上，其名称和类型分别是用户 id（int 类型），商品 id（int 类型），商品类目 id（int 类型），行为类型（varchar(50)类型），行为时间（int 类型），通过查询发现此表中共有 100150806 条记录。

3. 数据清洗

2.1. 数据去重

通过查询发现 userbehavior 中包含 49 条重复记录，取出其中 1000000 条不重复的记录插入到新建立的表 userbehavior2 中，此时已达到去重目的。

2.2. 缺失值处理

通过查询发现 userbehavior2 表中不含缺失值，对此不做处理。

2.3. 异常值处理

原数据集中包含的日期介于 2017 年 11 月 25 日至 2017 年 12 月 3 日，通过查询发现 userbehavior2 表中出现了不在此时间段的数据，将其视为异常数据，进行删除处理。

2.4. 一致性处理

数据集中行为时间这一字段为 unixtime 时间戳格式，需要利用 from_unixtime 函数对其进行转换，并向 userbehavior2 表中添加 v_date，v_time，v_hour 三个字段，分别表示行为所发生的日期，时间，小时。

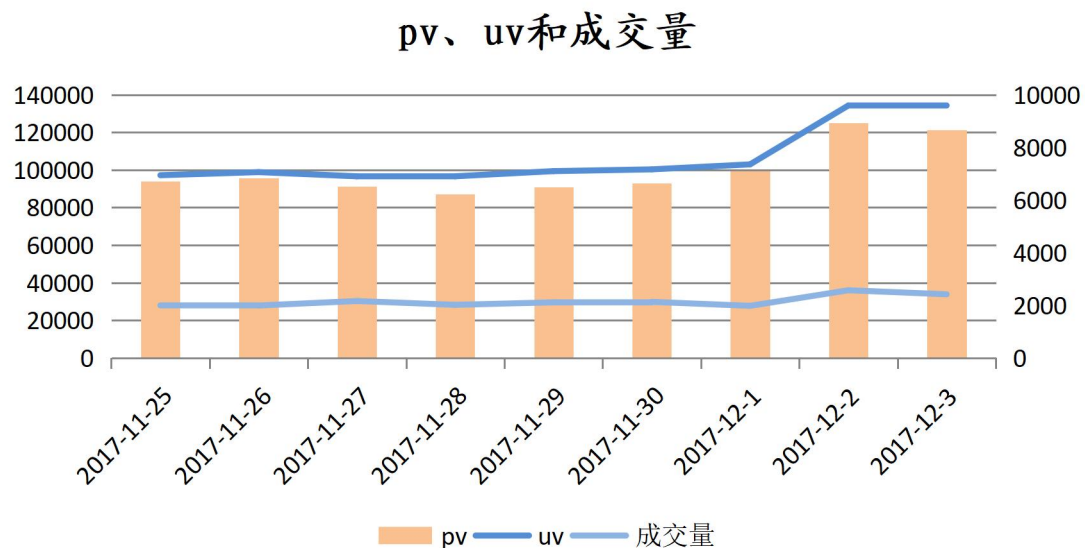
4. 模型构建与可视化分析

3.1. 用户行为分析

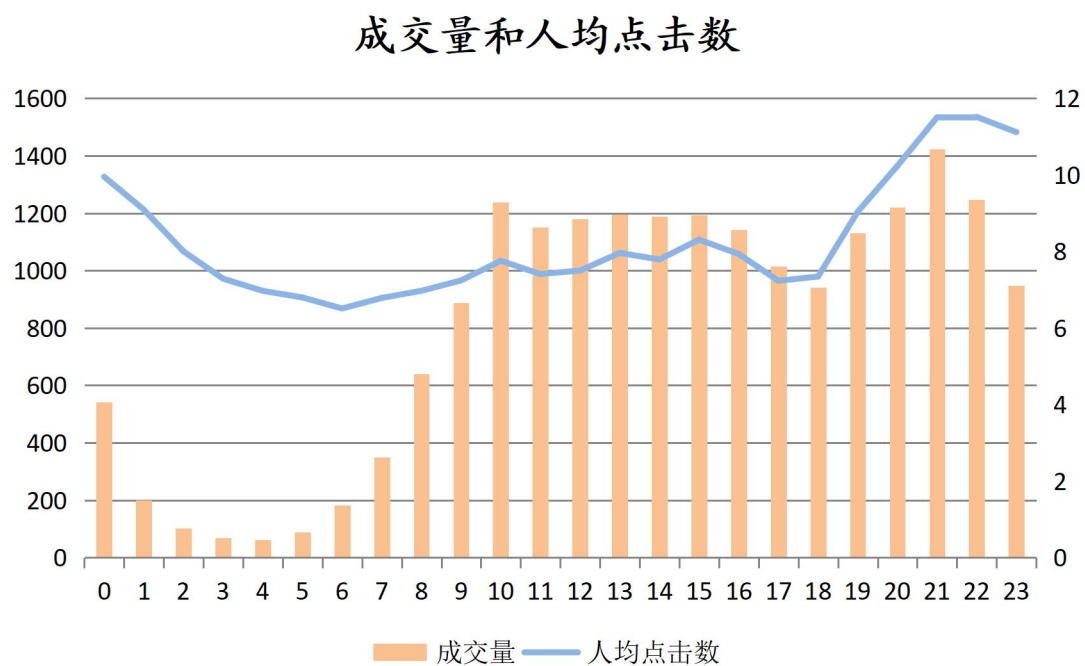
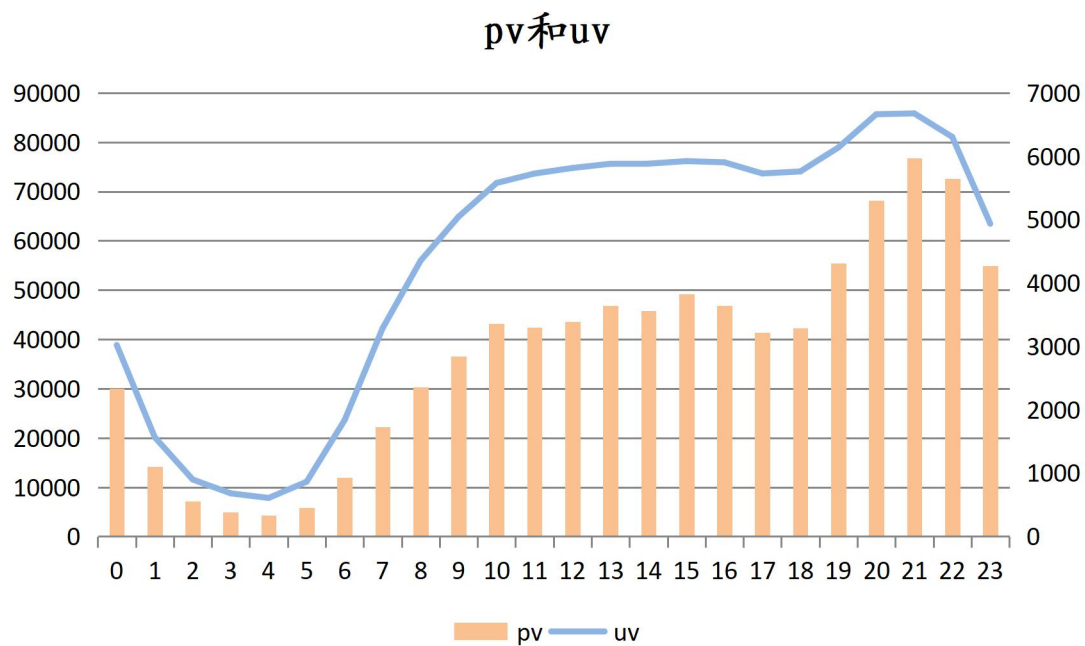
基于 userbehavior2 表查询独立访客量 uv，商品总数量，商品类型总数，行为类型总数，天数指标进行总览，结果显示 uv 为 9768 人，商品总数量为 392638，商品类型总数为 5651，行为类型总数为 4，包括点击、加购、收藏、购买，天数为 9 天。

查询 9 天内 uv，pv，人均浏览次数，成交量，结果显示 uv 为 9768 人，pv 为 897288 次，人均浏览次数为 9186 次，成交量为 19340。

查询日均 uv，pv，人均浏览次数，成交量，将结果可视化，由柱线图可知，日均 uv，pv，成交量自 12 月 2 日起大幅增加，推测是由于双十二活动预热导致。



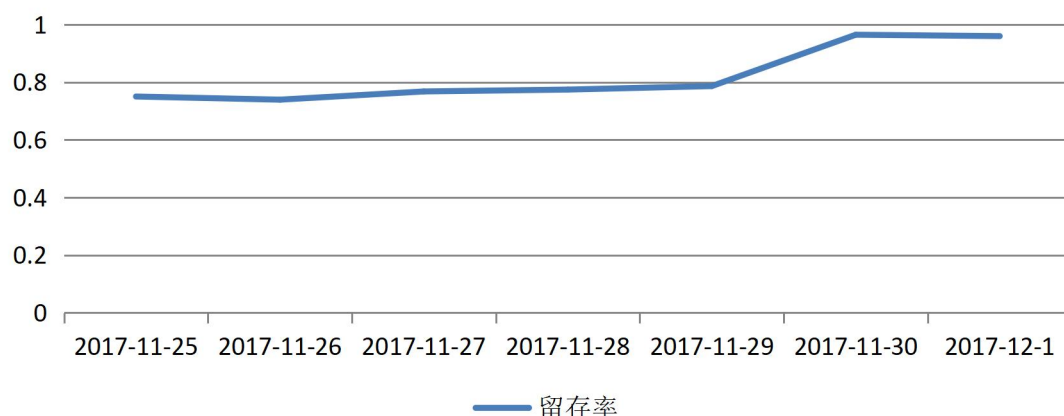
查询日均 uv，pv，人均浏览次数，成交量，将结果可视化，由柱线图可知，uv 和成交量在中午 12 时和晚上 21-22 时达到高峰，是用户一天内最活跃的时刻。商家应该集中资源，在这两个时间段进行引流与营销活动。



3.2. 用户留存分析

本项目没有给出判断是否为新老用户的依据，所以按照一般模型，计算每个日期对应的用户留存情况，以 pv 作为留存标准。以次日留存为例，做出可视化图表，由折线图可以看出，12 月 1 日的留存率明显提升，原因可能为发布了一些促进留存的运营活动。

次日留存率



3.3. 行为环节分析（漏斗模型）

对用户行为路径进行分析，用户购买行为遵循点击-加购/收藏-下单的连续过程，适用于漏斗模型，研究用户群体在此过程中的转化、流失情况，提出改善转化的建议。查询不同行为类型的独立用户数量及操作数，结果显示 9714 个用户进行了 897288 次点击，3601 个用户进行了 27316 次收藏，7338 个用户进行了 55547 次加购，6589 个用户进行了 19340 次购买。将收藏、加购看作一个环节，计算各环节的转化率：

- 环节转化率=本环节发生数/上一环节发生数
- 整体转化率=本环节发生数/第一环节发生数

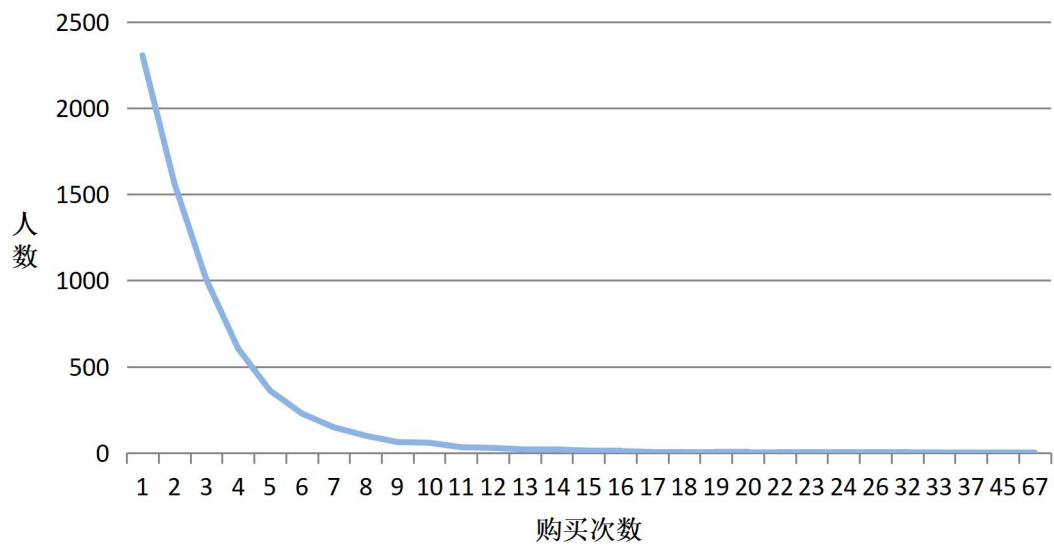
行为类型	独立用户数	点击	收藏	加购	成交
buy	6589	0	0	0	19340
cart	7338	0	0	55547	0
fav	3601	0	27316	0	0
pview	9714	897288	0	0	0

业务流程	次数	环节转化率	整体转化率
点击	897288	100.00%	100.00%
收藏、加购	82863	9.23%	9.23%
购买	19340	23.34%	2.16%

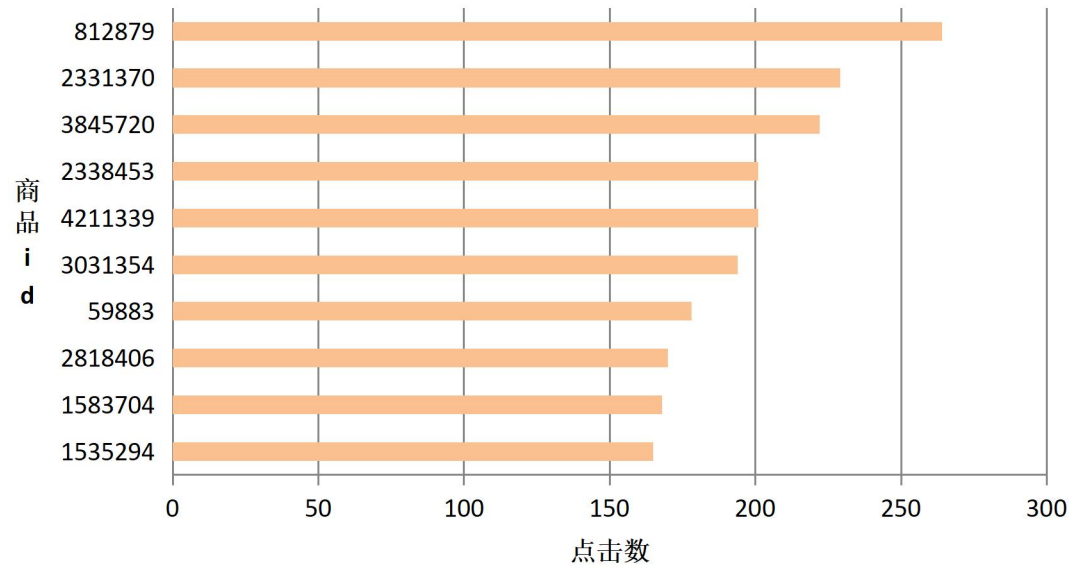
如上表可知，点击-收藏、加购的转化率为 9.23%，这意味着用户在进行点击后并没有进一步了解产品的欲望，用户流失主要发生在点击-收藏、加购这一环节。猜测原因可能为用户再点击后，进行收藏、加购的过程过于复杂，导致不愿意进行此操作，从而可以通过一些方法如简化收藏、加购流程，修改 APP 页面使得收藏、加购操作变得更加便捷减少用户流失。也有可能存在其他原因如宣传价格和加购价格不符，宣传图 and 实际不符导致的，具体原因需要更进一步的数据分析。

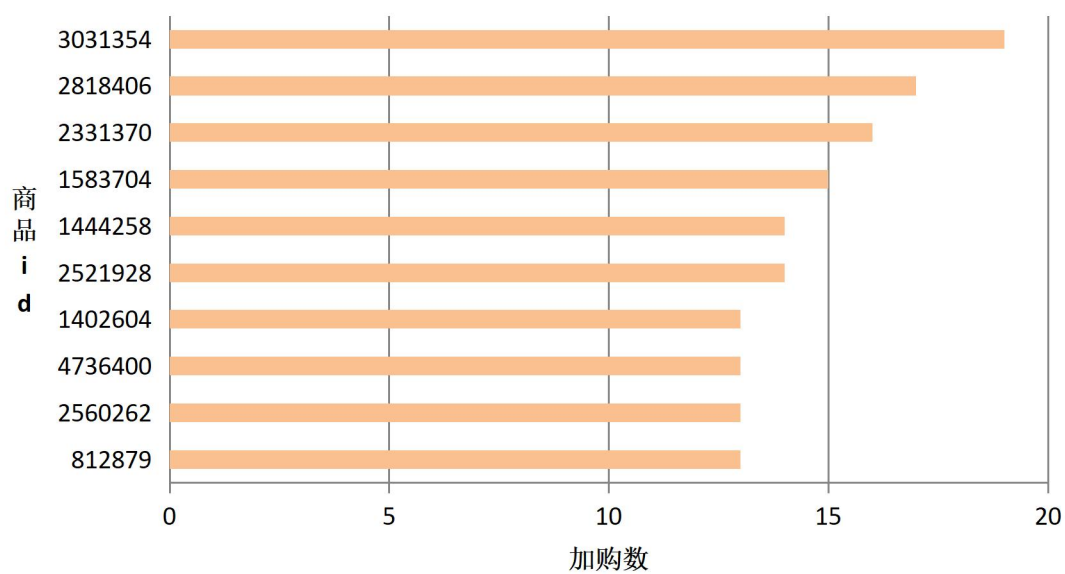
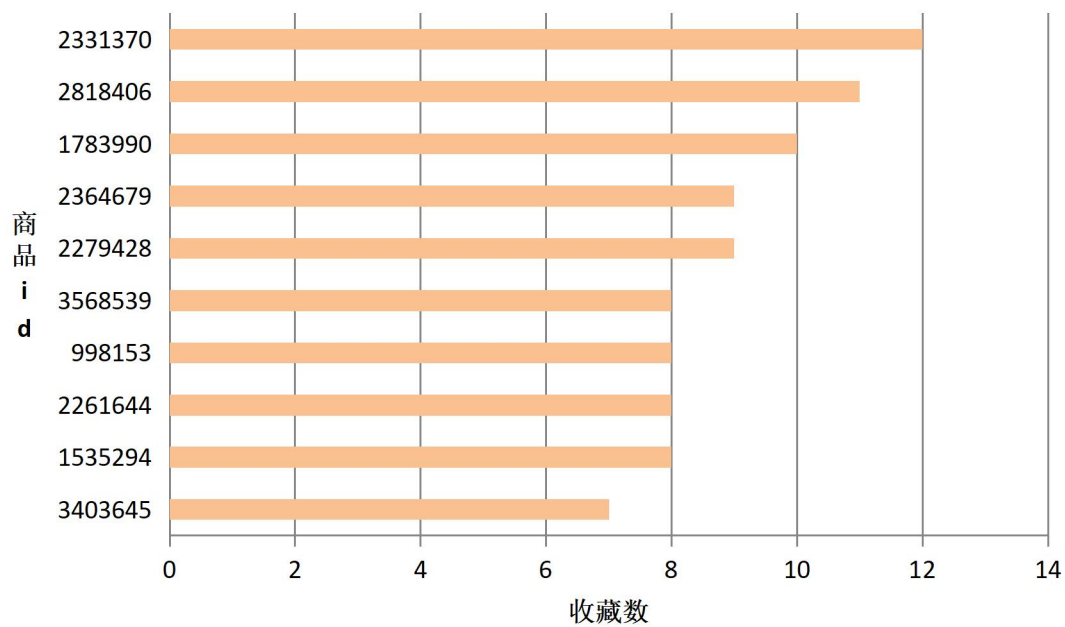
3.4. 商品维度分析

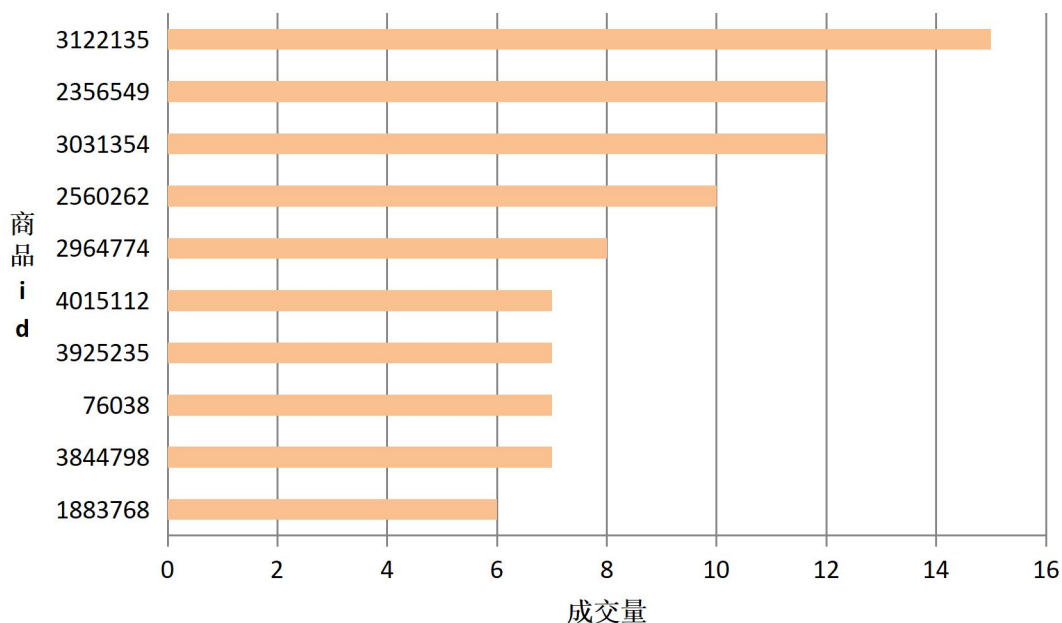
查询不同购买次数的用户数，将结果可视化，由折线图可以发现，购买次数为一次的用户占比 15.27%，购买次数为两次及以上的占比 84.73%，复购率较高。实际复购率要区分用户首单和首单转化等数据，此处无法判断真实复购情况。



分别查询点击、收藏、加购、购买的前十名商品，并将结果可视化，由条形图可知，商品 id 为 3031354 的商品位于点击、加购、购买前 10 名榜单。商品 id 为 2331370、2818406 的商品均位于点击、收藏、加购前 10 名榜，却没有在购买前十名榜单内，说明商家在用户的商品喜爱偏好上没有做到最好匹配。







3.5. 类目维度分析

查询商品类目的渗透率，结果显示渗透率最高的商品类目为 1464116，购买过 146116 商品类目的用户买过其他 423 类不同的商品类目，渗透率前五名的商品类目 id 为 146116，4756105，2735466，4145813，2885642，在营销过程中，通常选择渗透率高的类别作为重点营销类别。

3.6. 用户价值分析（RFM 模型）

用户价值分析基于 RFM 模型，找出高价值用户，对其进行精准营销，确定精细化营销策略，对不同价值的用户采用不同的运营策略来刺激消费。RFM 模型中 R 表示最近一次消费的时间间隔，F 表示某一段时间内的消费次数，M 表示某一段时间内的消费金额，由于数据集中没有消费金额，所以只根据 R 和 F 对用户进行价值分析。

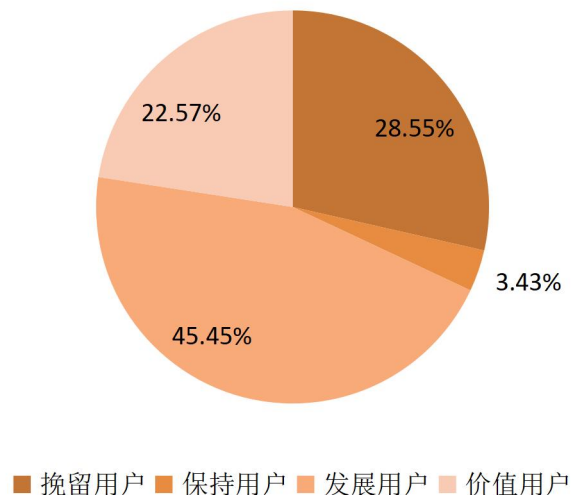
第一步，创建视图 RFM，用来存放用户最近一次消费的时间间隔和 2017 年 11 月 25 日 -2017 年 12 月 3 日的消费次数。通过查询统计各个时间间隔的用户数和各购买次数的用户数。

第二步，给 R、F 进行打分，计算每个用户的 R、F 值得分，如果用户的 R、F 值大于所有用户 R、F 值的平均值，则视为“高”，否则视为“低”。

分数	R 值	F 值
1	8	1-3
2	6, 7	4-7
3	4, 5	8-16
4	2, 3	17-26
5	0, 1	32-67

按照划分依据可知，可以将用户分为四种类型，分别为价值用户（R 值高，F 值高），

发展用户（R 值高，F 值低），保持用户（R 值低，F 值高），挽留用户（R 值低，F 值低），统计各类用户数，可以发现发展用户>挽留用户>价值用户>保持用户，比较价值用户数和保持用户数发现价值用户数的 R 值为高，对应最近购物时间间隔小的用户，价值用户的数量大于保持用户的数量原因可能在于双 12 活动预热导致最近购买的用户数量大幅增加。



在开展营销活动之前，对用户进行价值分析，可以针对不同用户进行精细化服务，采用不同的营销策略。

5. 营销策略

由不同的业务需求可以提出不同的运营和营销策略。

获取客户：邀请好友点击领取优惠券，或通过拼团、砍价增加平台用户数，同时加大商品广告投放。

激活用户：当平台中新进入的用户数较多、用户池大，但是使用率却很低的情况下，需要完成用户激活。通过分析本数据集我们发现用户流失主要发生在点击-加购或收藏这一环节。我们猜测可能是因为用户点击加入购物车或收藏的过程太过于复杂，导致用户不愿继续进行下一步操作导致的，这时我们可以对用户加入购物车或收藏的过程提出一些优化建议，如首页直接推荐商品，不放无效信息，不需要用户点击多次才到商品页面；不设置购物车，点击商品后直接购买支付，减少用户犹豫时间；先付款后拼团，将支付流程提前，系统自动将所有用户以团购价自动拼团；通过游戏娱乐，给用户发放奖励来唤醒用户，例如：砍一刀减价等。

提高留存：要想提高留存就是要培养用户的习惯，这里可采取的措施有：对所有商品包邮，让用户购物中习惯免邮费，给购买次数多且金额大的用户打折，让用户习惯会员折扣，从而不再去别的平台买东西。设置新用户奖励和一定复购次数用户折扣，让用户能从获取平台优惠券或者积分，减少付款金额。

增加收入：将收入分为服务收入和广告收入，找损失潜在收益的地方，分析用户关键环节放弃的原因，细化解决相应问题。例如，在服务收入上，根据 RFM 分析方法确定的用户类型，对不同的用户采取不同的措施，从而提高复购率、成交量。对价值用户：提供 VIP 服务，当这类用户再次在平台购买时，给予一定的 VIP 折扣：满减或者打折；对发展用户：其购买次数低，这时候要想办法提高其购买频率：通过短信方式，适当地赠送优惠券；对保

持用户：这类用户购买次数高，但是购买时间间隔太长，属于一段时间没来的忠实用户，这时候应该主动与用户保持联系，提高复购率；对挽留用户：这类用户购买次数少且时间间隔长，属于即将流失的用户，这时候需要主动联系用户，弄清楚原因，想办法挽留。而在广告收入上，根据之前分析的 Top10 部分，找出用户偏好，投放需求度高的商品广告和营销活动，尽可能放在流量多的时间段进行推荐和宣传。