

# 基于 SQL 的淘宝用户行为分析

## 分析背景

淘宝网，是全球最受欢迎的网购零售平台之一，拥有近 5 亿的注册用户数，每天有超过 6 千万的固定访客，同时每天的在线商品数已经超过了 8 亿件，平均每分钟售出 4.8 万件商品。

作为电商行业的标杆，淘宝网的用户行为数据，一定程度上反映了用户的购买行为规律。本项目利用 MySQL 对淘宝用户行为数据进行分析，探索不同用户的消费习惯，结合店铺营销策略，以实现精准化运营，减少运营成本，增加店铺营收。

## 分析目标与思路

**探索目标：**每日不同时段流量走势，质量如何?用户行为转化情况怎样?如何提高留存、增加复购?如何判断高价值用户，针对不同用户如何进行个性化营销?

### 分析思路：

- 流量指标分析：**对比每日每小时的 PV（访问量）、UV（访客数）、平均访问量（PV/UV），在不同时间发布不同的营销活动，寻找更优质的渠道。
- 行为转化分析：**统计用户不同行为的转化情况，简化交易流程，提高转化率。
- 产品贡献定量分析：**根据产品贡献，调整产品结构，选定营销主题。
- 用户价值分析：**对用户进行价值分层，针对不同层级的用户制定不同的营销策略。

## 数据准备

### 1. 数据获取

数据集来源于淘宝 APP 移动端，2014 年 11 月 18 日至 2014 年 12 月 18 日的用户行为数据。

### 2. 数据理解

表中有 7 个字段：用户 ID、商品 ID、商品类目 ID、用户行为类型、用户所在位置、用户行为时间、金额，一共有 12256905 条数据。限于电脑性能，仅提取前 1048575 条记录用作分析。

字段名	字段描述	数据类型
user_id	用户ID	INT
item_id	商品ID	INT
item_category	商品类目ID	INT
behavior_type	用户行为类别	VARCHAR(10)
user_geohash	用户地理位置	VARCHAR(10)
times	行为时间	DATETIME
amount	金额	DECIMAL(5,2)

**pv:** 点击，用户访问商品详情页。必须是商品详情页的流量，淘宝网首页、搜索结果页等页面的点击流量不算在内。

**fav:** 收藏，用户将商品加入收藏夹，便于下次访问。收藏夹中的商品不可直接购买，必须再次发生“点击”行为后，在商品详情页方可购买。

**cart:** 加购，用户将商品添加到购物车。用户可从购物车直接购买商品，无须再次产生“点击”行为。

**buy:** 下单，用户点击提交订单购买商品。

3. 数据清洗

**3.1 缺失值处理:** 查询结果中发现 `user_geohash` 字段存在缺失值，缺失比例超过 50%。缺失严重的数据在后续分析中意义不大，可做删除处理。

**3.2 异常值检查:**检查发现该数据集中不存在异常值。

**3.3 重复记录处理:**该数据集有 86478 条重复记录，筛选不重复的记录共 962097 行。

**3.4 字段处理:**

根据 `times` 字段增加计算字段用户行为日期、周和小时，排除后续分析不需要的 `user_geohash` 字段，并将筛选后的结果保存到新表。

清洗后的数据集共 962097 条记录，展示了 2014-11-18 至 2014-12-18 这一个月，8477 位用户对 7095 个类目下 701609 个商品产生的 902050 次点击，29016 次加购，20930 次收藏，10101 次购买行为。

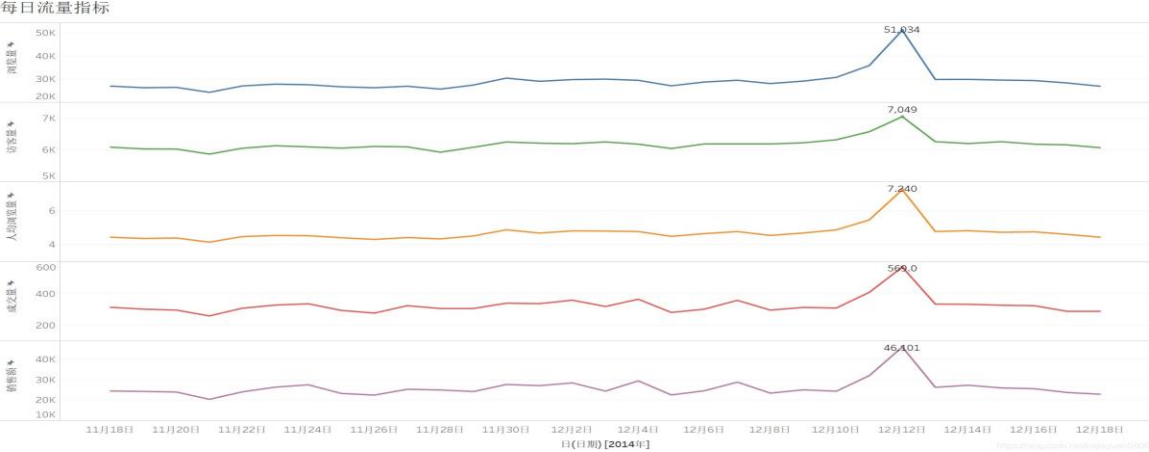
数据分析

1.流量指标分析

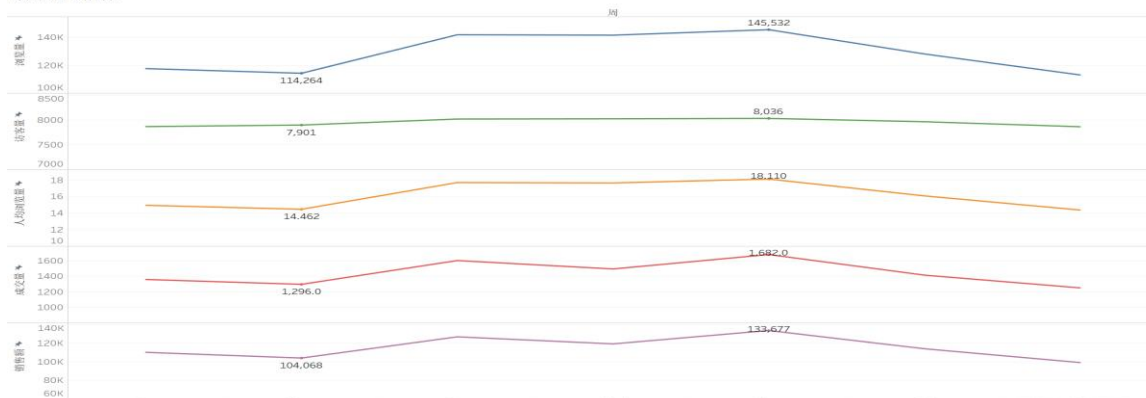
a. 浏览量 PV: 页面浏览量或点击量 **Page View**，用户每打开一个网页就记录 1 次 PV，用户多次访问同一页面 PV 累计多次。

b. 访客数 UV: 访问页面的独立访客 **Unique Visitor**，用户访问该页面的一台电脑客户端为一个访客，以浏览器的 `cookie` 为依据，同一个用户多次访问只计算一个 UV。

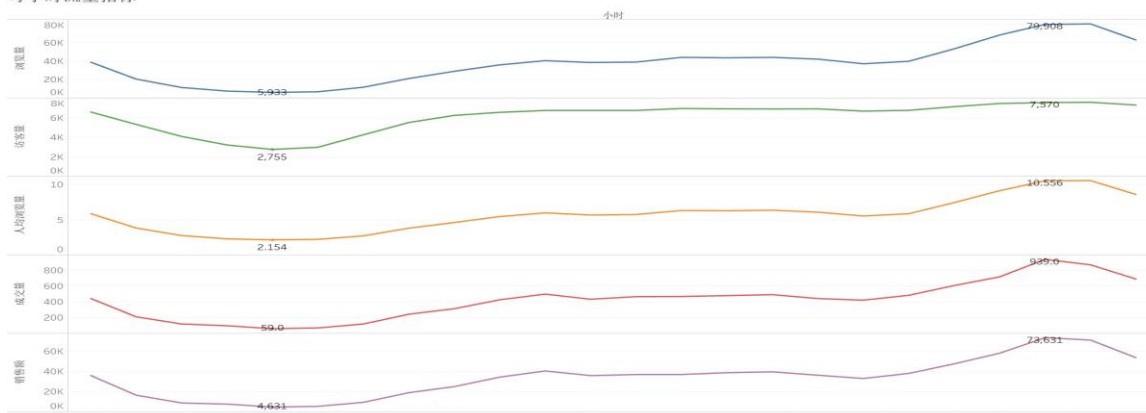
c. 人均浏览量 PV/UV: 平均每个独立访客的浏览次数，体现网站对用户的吸引程度。



每周流量指标



每小时流量指标



查询结果中发现浏览量 PV、访客量 UV 和人均浏览量的总体变化趋势是相同的。11 月 18-29 日，PV 在 25000-28000 之间，UV 在 5800-6200 之间，人均浏览量在 4.1-4.6 之间，成交量在 260-320 之间，销售额在 20000-28000 之间；11 月 30 日-12 月 10 日，PV 在 27000-31000 之间，UV 在 6000-6400 之间，人均浏览量在 4.4-4.9 之间，成交量在 280-370 之间，销售额在 22000-30000 之间；12 月 11 日、12 日因双 12 的缘故，PV 分别达到了 35851、51034，UV 分别为 6576、7049，人均浏览量分别为 5.5、7.2，成交量分别为 409、569，销售额分别为 32022.71、46100.93；而在双 12 之后，PV 也稳定在 28000-30000 之间，UV 在 6000-6200 之间，人均浏览量在 4.6-4.8 之间，成交量在 280-340 之间，销售额在 22000-28000 之间，远远优于 11 月 29 日之前的数据，说明双 12 的营销活动对用户行为有显著影响。

同时，在 11 月 21、28 日，12 月 5 日，浏览量 PV、访客量 UV、人均浏览量、成交量和销售额都出现了不同程度的下滑。细查发现以上三天都是周五，观察流量的周期性变化中发现，每周二到周四数据比较稳定，周五到周一会出现小幅度下滑，这跟大部分用户的工作和生活习惯息息相关。

从每天的时间段来看，21 点和 22 点是用户活跃的高峰时间，人均浏览量在 10 次以上，22 点之后一路下降，一直到凌晨 4 点降到最低值，4 点到 10 点渐渐回温，10 点到 18 点基本稳定，18 点之后开始一路攀升，商家可以集中资源，在用户活跃度较高的时间段采取一些引流手段。

## 2.行为转化分析

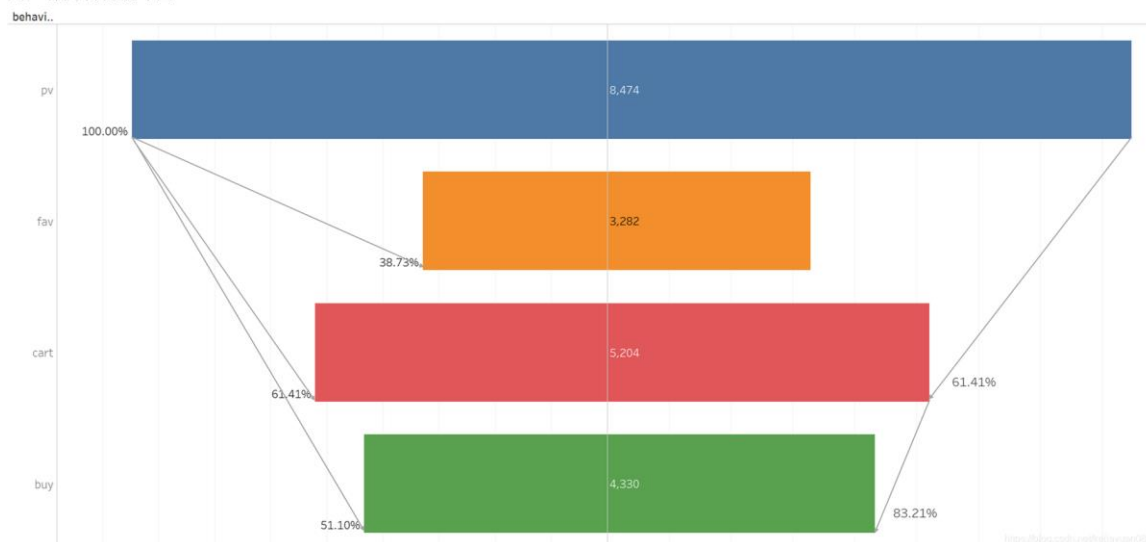
浏览人数：点击产品详情页的用户数

收藏人数：收藏产品产品详情页的用户数

加购人数：将产品添加到购物车的用户数

购买人数：下单的用户数

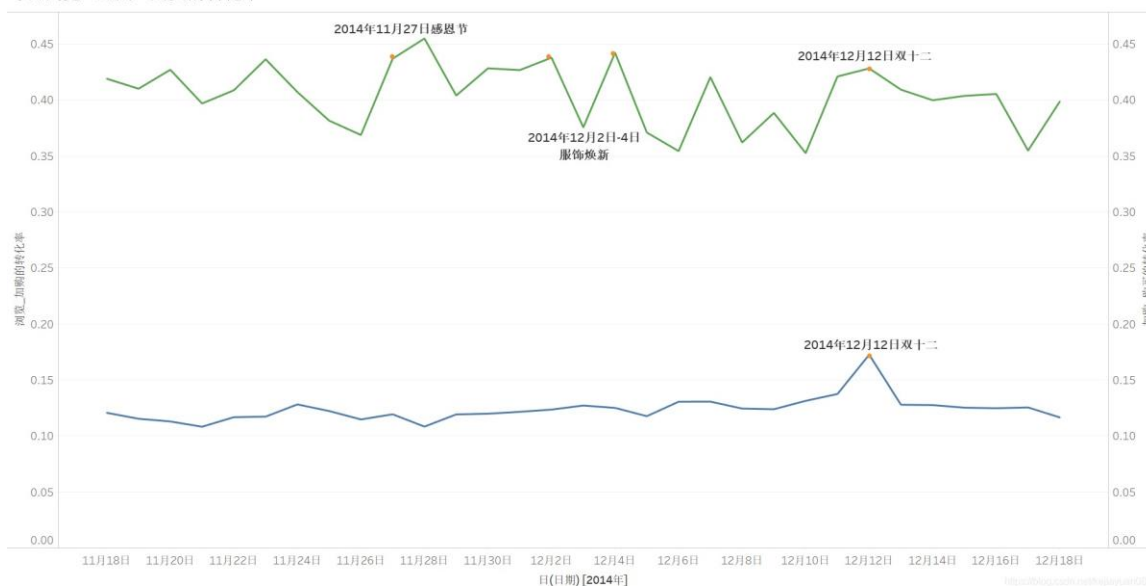
用户行为转化漏斗图



在所有用户行为数据中浏览产品详情页后产生购买行为的用户只有 4330 人，占浏览用户的 51%，也就是说，还有 49% 的用户行为是没有转化为成交的，用户在浏览商品详情页后出现了大量的流失。那么从浏览到购买，每个环节的转化率是多少？用户主要是在哪个环节流失的呢？

对比发现，从浏览到收藏的转化率仅为 38.73%，而浏览到加购的转化率为 61.41%。因为收藏和加购行为之间并没有直接的先后关系，且加入收藏后并没有可以下单的入口，如果需要购买必须重新点击商品进入详情页才能下单，所以用户更愿意将感兴趣或有购买意向的商品添加到购物车。

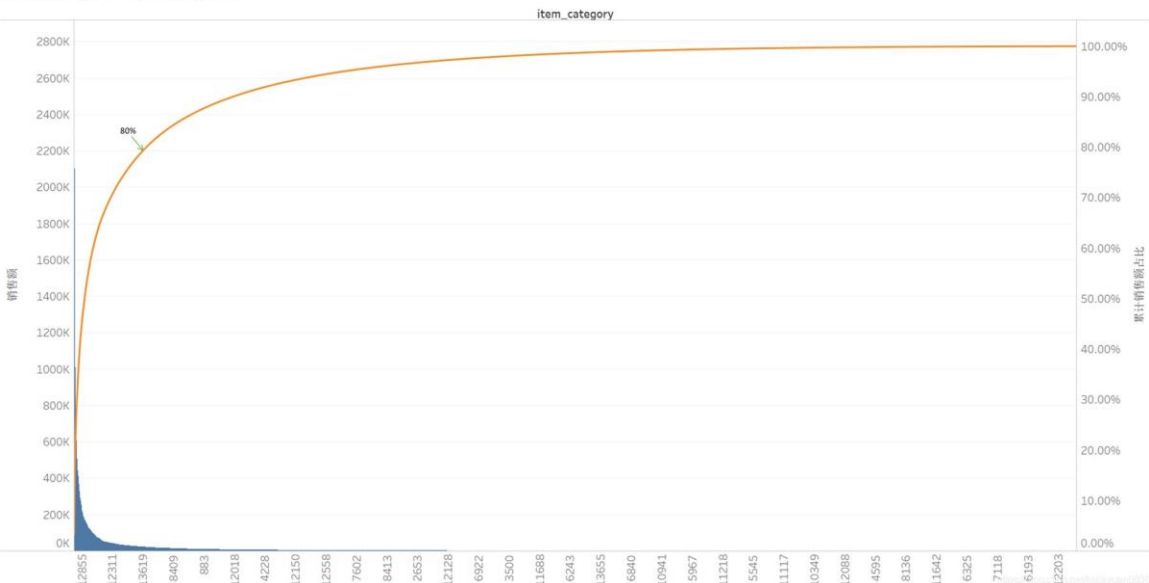
每日浏览—加购—购买的转化率



分析每天浏览-加购-购买的转化率情况发现，浏览-加购的转化率变化并不明显，双十二之前稳定在 10%-13%之间，仅在双十二当天达到 17.26%，且在双十二之后转化率明显下降。而加购-购买的转化率波动较大，在感恩节、服饰焕新和双十二活动前后分别达到了 45.48%、44.21%、42.81%，而双十二之后的数据表现也明显优于双十二之前，说明部分商品在双十二之后依然有返场活动，促进用户消费。

### 3.产品贡献定量分析

帕累托分析：产品类别-销售额



在 2066 个商品类目中，销售额排名前 623 个类目贡献了 646009.09 的销售额，可集中人力和财力优化这些类目的商品及渠道。

### 4.用户价值分析

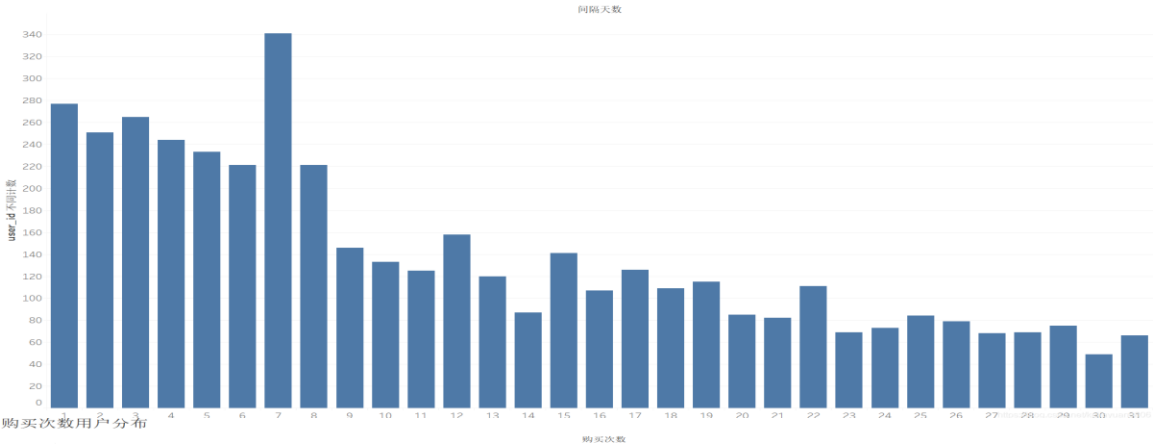
通过对用户价值的细分，进行差异化的精细运营，从而提升运营效率和用户体验。RFM 模型是衡量客户价值和客户创利能力的重要工具，通过客户的消费时间间隔（Recency）、消费频率（Frequency）以及消费金额（Monetary）三项指标来描述客户的价值状况。

最近一次消费时间间隔（R）：上一次消费离得越近，也就是 R 的值越小，用户价值越高。

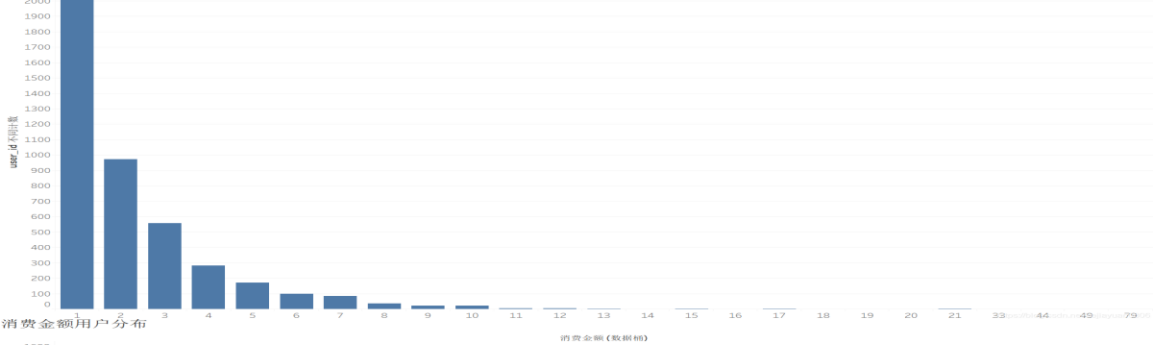
消费频率（F）：购买频率越高，也就是 F 的值越大，用户价值越高。

消费金额（M）：消费金额越高，也就是 M 的值越大，用户价值越高。

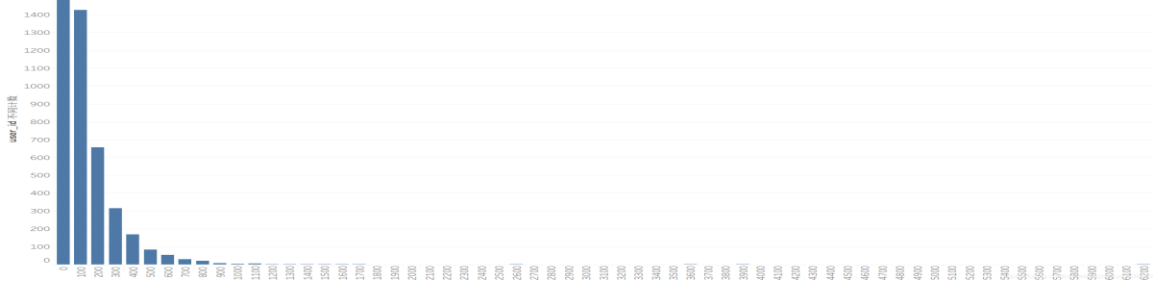
间隔天数用户分布



购买次数用户分布



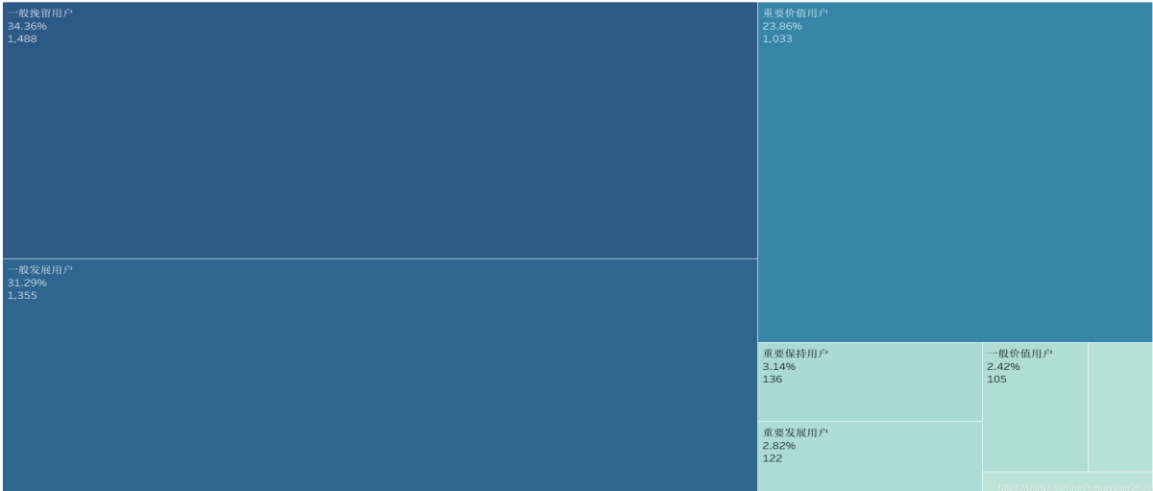
消费金额用户分布



根据用户数量分布，制定以下评分标准：

分值	间隔天数	购买次数	消费金额
1	25-31天	1次	[0,100)
2	19-24天	2次	[100,200)
3	13-18天	3次	[200,300)
4	7-12天	4次	[300,400)
5	1-6天	5次及以上	400及以上

用户价值细分



对用户进行价值分层后，针对不同价值类型的用户实行不同的营销策略。

一般挽留用户占比 34.36%，占比最高，其交易时间间隔长，交易频率低，消费金额低，存在流失风险，可以及时与用户取得联系，明确流失原因或了解用户需求，想办法挽回用户。

一般发展用户占比 31.29%，占比排名第二，其交易时间间隔短，但消费频率和消费金额都很低，可以利用推荐系统推荐其平时浏览的同类商品，或与此类客户有相同购买属性人群购买的商品，发送满减优惠券等，避免用户流失。

重要价值客户占比 23.86%，其交易时间间隔短，消费频率高，消费金额高，应加强交流与互动，深入了解用户需求，提供个性化服务，增加用户粘性。可以对这类用户提供 **VIP** 服务机制，提升用户体验与忠诚度。

## 总结和建议

1. 根据流量数据指标分析，每天的 18-22 时是用户活跃高峰期，可集中资源在该时间段进行引流与营销活动，如推送爆款及畅销商品，同时发送商家折扣优惠、直播带货及促销活动消息，提高商品购买率。
2. 从用户行为转化漏斗分析，浏览到加购的转化率远远高于浏览到收藏的转化率，且加购到购买的转化率高达 83.21%。交易流程越多，用户流失的可能性就越大，瞬间购买欲望的涌现，往往时效性较短，每多一步流程，就多一份流失率。所以删除不必要的跳转界面，让交易的完成时间更短，不要给用户过多的犹豫时间。例如优惠券的选择，还需要繁琐的领取步骤，满足优惠条件默认领取相应最大的优惠，并在最后付款界面突出优惠券带来的满减信息，促使用户的成交花费时间更短。关于购物车，可以增加一个清理机制，比如加入购物车以后多少天不成交就清除，每隔一段时间就提醒用户购物车内还有什么商品未下单，什么时间商品将被清除。收藏和购物车的功能有一定的同质性，关于这一点，淘宝没有做明显的区分，导致购物车只是简便版的收藏。反观同电商领域的拼多多，没有设置购物车功能，并且成交环节缩短至 4 步。
3. 根据用户消费偏好分析，将流量更多地聚焦于畅销的商品，通过明星效应/网红推荐，结合主题活动和节日，打造爆款的产品，并利用爆款产品带动整体商品的销售。

4. 根据 RFM 用户价值分类结果，一般挽留用户存在流失风险，可适当给予折扣或捆绑销售政策，通过一些补贴优惠，培养用户的购买习惯。一般发展用户存在价值潜力，要想办法主动联系客户，如短信、邮件、push 提醒等方法，进行好物推荐等，提高用户的复购率。重要价值客户其交易时间间隔短，消费频率高，消费金额高，应加强交流与互动，比如优化个人信息界面，加入成长值系统，到达一定等级，享受一些特权或优惠，可参考腾讯游戏的心悦会员制度。