

# 淘宝用户行为分析

## 一、数据来源及含义

### 1、数据来源

阿里云: <https://tianchi.aliyun.com/dataset/dataDetail?dataId=72423>

文档 CSDN 网址: [https://blog.csdn.net/qq\\_33923785/article/details/123453999](https://blog.csdn.net/qq_33923785/article/details/123453999)

原始数据有 1048575 条, 数据量适中, 方便使用 Mysql 进行操作。

### 2、字段含义

MYSQL 数据库展示: 创建表并导入数据:

```
CREATE TABLE user_behavior
```

```
(uid VARCHAR(20), item_id VARCHAR(20), item_cate_id VARCHAR(20), b_type VARCHAR(20), `time` VARCHAR(20) );
```

## (将字段类型设置成字符串类型是为了在数据库导入 EXCEL 文件时避免数据丢失)

(提示: 在导入数据库数据时, 可以将获取的 .CSV 数据集在 excel 中打开, 选择设置“数据-分列-文本格式”, 保存后在数据库中选择“导入本地加载的 CSV 数据”。这样无论是数字还是字符都可以无损导入到数据库中。导入后在进行字段类型修改)

Field	Type	Collation	Null
uid	varchar(20)	utf8_general_ci	YES
item_id	varchar(20)	utf8_general_ci	YES
item_cate_id	varchar(20)	utf8_general_ci	YES
b_type	varchar(20)	utf8_general_ci	YES
time	datetime	(NULL) CSDN @qq_33923785	YES

解释: uid: 用户 id item\_id: 商品 Id item\_cate\_id: 商品类别 id

b\_type: 用户行为类型 time: 时间戳

其中 b\_type 包括: 1 (点击) 2 (收藏) 3 (加入购物车) 4 (购买)

## 二、数据清洗

### 1、添加辅助字段

①将原有数据中 ‘1’、‘2’、‘3’、‘4’ 修改为 ‘点击’、‘收藏’、‘加购’、‘购买’

```
UPDATE user_behavior SET b_type='点击' WHERE b_type='1';
```

```
UPDATE user_behavior SET b_type='加购' WHERE b_type='2';
```

```
UPDATE user_behavior SET b_type='收藏' WHERE b_type='3';
```

```
UPDATE user_behavior SET b_type='购买' WHERE b_type='4';
```

②: 修改字段类型、增加辅助字段

##更改字段 time 格式

```
ALTER TABLE user_behavior MODIFY `time` DATETIME; ##提取日期数据
```

```
ALTER TABLE user_behavior ADD COLUMN dates CHAR(10) NULL;
```

```
UPDATE user_behavior SET dates =SUBSTR(`time`,1,10); ##提取小时
```

ALTER TABLE user\_behavioror ADD COLUMN `hour` CHAR(10) NULL;

UPDATE user\_behavioror SET `hour` = SUBSTR(`time`,12,2);

进行上述处理后结果如下:

uid	item_id	item_cate_id	b_type	time	dates	hour
99989881	4386604	874	点击	2014-12-02 18:00:00	2014-12-02	18
99989881	202234711	2993	点击	2014-12-09 21:00:00	2014-12-09	21
99989881	152460728	3472	点击	2014-11-30 23:00:00	2014-11-30	23
99989881	402254968	2993	点击	2014-12-09 21:00:00	2014-12-09	21
99989881	318841119	2993	点击	2014-12-15 22:00:00	2014-12-15	22
99989881	160680305	2993	点击	2014-12-05 01:00:00	2014-12-05	01
99989881	318841119	2993	点击	2014-12-09 20:00:00	2014-12-09	20
99989881	289148767	3472	点击	2014-11-30 23:00:00	2014-11-30	23
99989881	254028479	874	点击	2014-12-10 11:00:00	2014-12-10	11

## 2、异常值

### ①查看是否有异常值

##查看是否有异常值

```
SELECT MAX(`time`), MIN(`time`), MAX(dates), MIN(dates)
FROM user_behavioror;
```

MAX(`time`)	MIN(`time`)	MAX(dates)	MIN(dates)
2014-12-18 23:00:00	2014-11-18 00:00:00	2014-12-18	2014-11-18

此分析针对‘2014-11-18-2014-12-18’数据，数据符合预期。

### ②查看数据集

SELECT DISTINCT \* FROM user\_behavioror; ##共有取出数据 971187 条

## 3、缺失值

```
SELECT COUNT(uid),COUNT(item_id),COUNT(item_cate_id),COUNT(b_type),COUNT(`time`)
FROM user_behavioror;
```

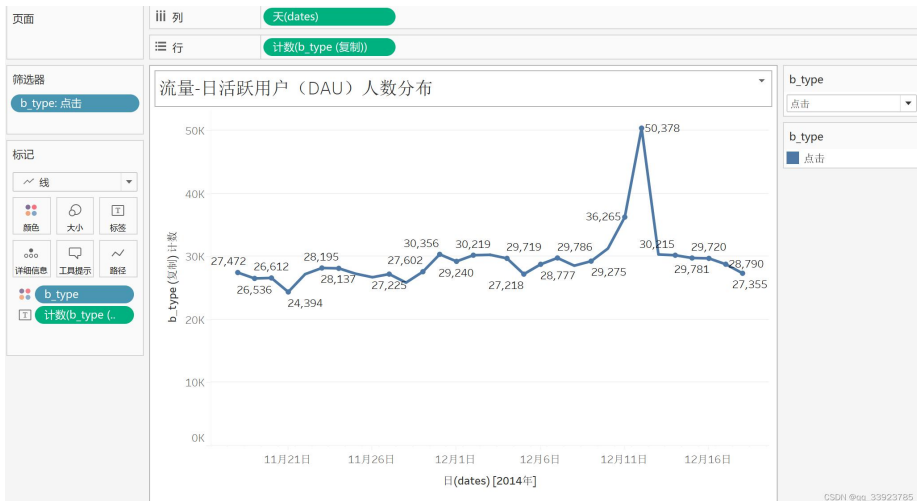
COUNT(uid)	COUNT(item_id)	COUNT(item_cate_id)	COUNT(b_type)	COUNT(`time`)
1048575	1048575	1048575	1048575	1048575

显示无缺失值

## 三、模型构建

### 1、用户流量分析

#### ① 日活跃用户（DAU）人数分布



2014 年 11 月 18 日-2017 年 12 月 18 日 活跃人数整体较为平稳，但在增长趋势，11 月 18 日-12 月 10 日波动趋势始终在 30K 上下，12 月 11 日临近淘宝“双十二”购物节，DAU 出现明显的上扬增长。12 月 11 日人口较 10 日增长 24%，12 日较 10 日增长约 70%。13 日 DAU 迅速下降到正常人数水平。

## ② PV、UV、PV/UV

```
SELECT COUNT(DISTINCT uid)AS 'UV',
(SELECT COUNT(*) FROM user_behavior WHERE b_type='点击')AS 'PV',
(SELECT COUNT(*) FROM user_behavior WHERE b_type = '点击')/(COUNT(DISTINCT uid) ) AS 'PV/UV'
FROM user_behavior;
```

UV	PV	PV/UV
8477	987911	116.5402

在 PV/UV 比上可以明显看出，双十二活动对整个月的点击、访客数的影响。因为淘宝双十二活动并不是只是只在当天进行，而是在前后半个月左右就开始预售活动，访客、点击数都会受到平台活动的影响。

注：UV：独立访客数 PV：页面流量次数 PV/UV：平均一个独立访问者浏览的访问量

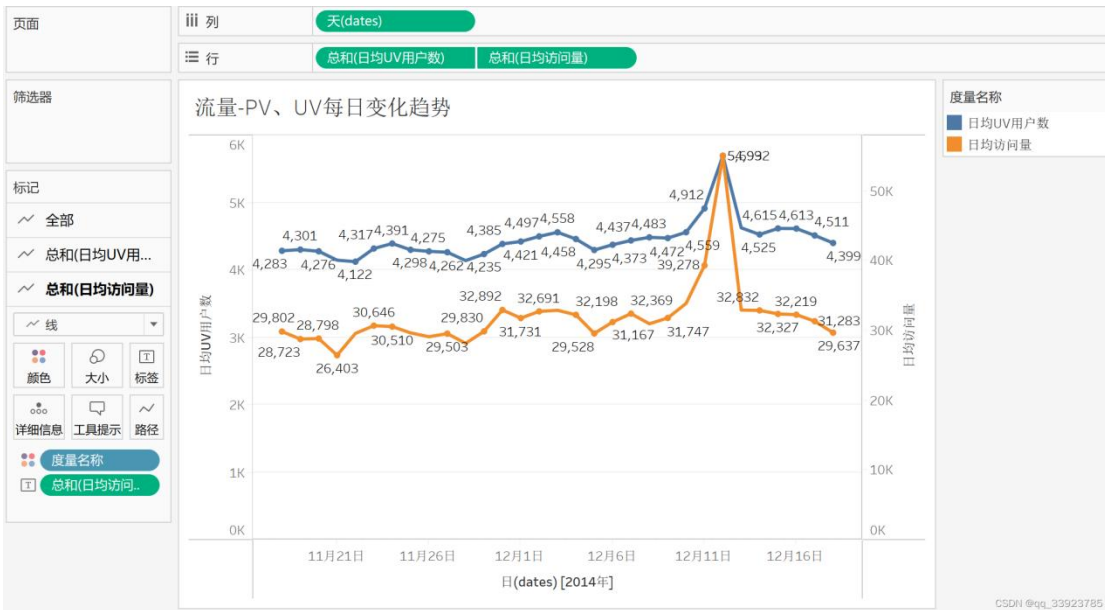
## ③ PV、UV 每日变化趋势

```
SELECT dates, SUM(CASE WHEN b_type='点击' THEN 1 ELSE 0 END)AS '日均访问量',
COUNT(DISTINCT uid)AS '日均UV用户'
FROM user_behavior GROUP BY dates ORDER BY dates;
```

dates	日均访问量	日均UV用户
2014-11-18	29802	4283
2014-11-19	28723	4301
2014-11-20	28798	4276
2014-11-21	26403	4144
2014-11-22	29538	4122
2014-11-23	30646	4317
2014-11-24	30510	4391
2014-11-25	29615	4298
2014-11-26	29060	4275
2014-11-27	29503	4262
2014-11-28	28119	4138
2014-11-29	29830	4235
2014-11-30	32892	4385
2014-12-01	31731	4421
2014-12-02	32691	4497
2014-12-03	32846	4558
2014-12-04	32198	4458
2014-12-05	29528	4295
2014-12-06	31167	4373
2014-12-07	32369	4437
2014-12-08	30913	4483
2014-12-09	31747	4472
2014-12-10	33834	4559
2014-12-11	39278	4912
2014-12-12	54992	5693
2014-12-13	32880	4624
2014-12-14	32832	4525
2014-12-15	32327	4615
2014-12-16	32219	4613
2014-12-17	31283	4511
2014-12-18	29637	4389

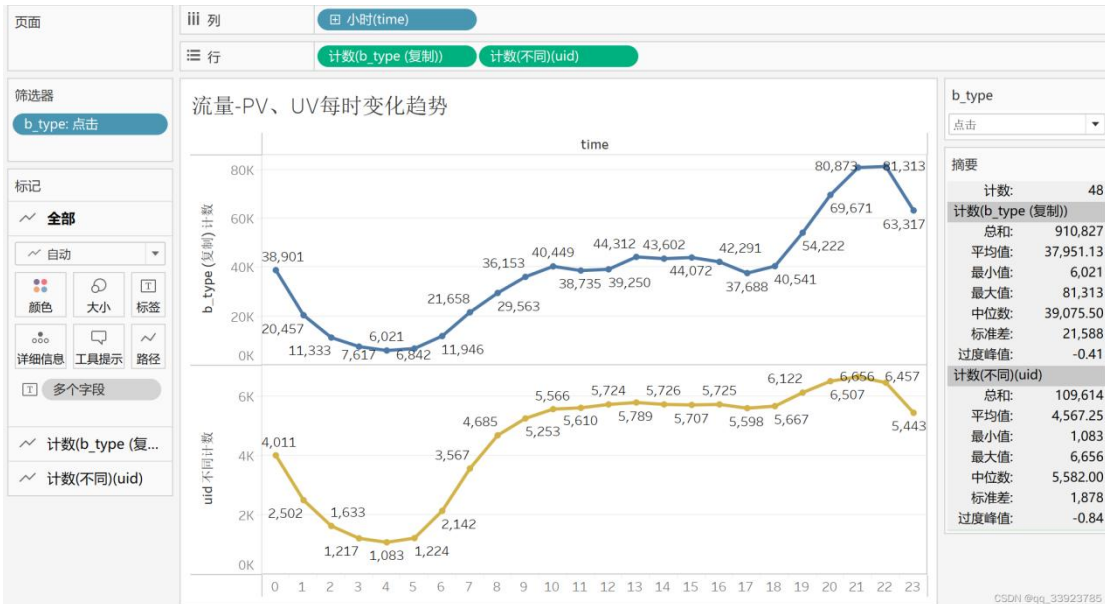
CSDN @qq\_334389

将上述查询所得表格导入 Tableau 中，绘制 PV、UV 每日变化趋势图。



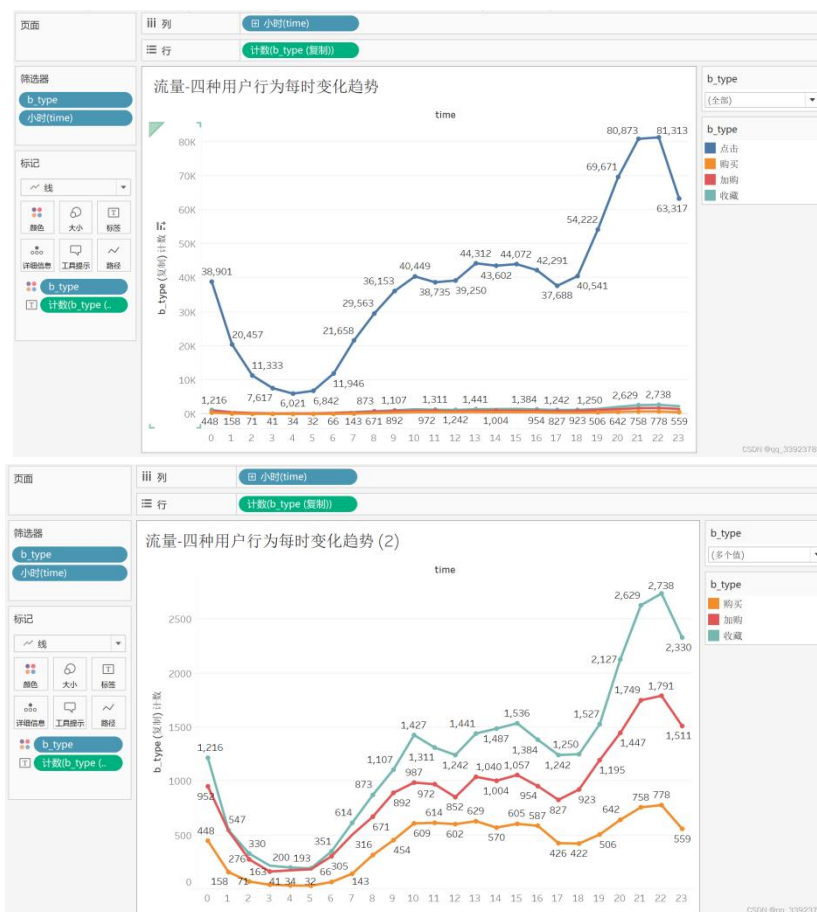
可以看出，每日的访问人数变化和访问次数变化基本上走势趋于一致。在 12 月 12 日，二者都达到了最高峰，并且相互重叠。12 月 12 日是淘宝系列产品的“双十二活动”日，在该日前后一个周期内，用户访问和用户消费都会呈现出一个明显的变化。并且在双十二单日变化最大。

#### ④ PV、UV 每时变化趋势



从图可以看出来，每天结合工作日考虑，人们一般都是早八晚六，或者早十晚七。在每天的下班时间点即 17-19 点左右，点击次数和访客数都会出现一些小幅上涨。

## ⑤ 四种用户行为每时变化趋



从上图可以看出：

- (1) 四种行为的分布趋势基本一致。
- (2) 用户在晚上 18 点到 23 点会频繁点击浏览网页，收藏与加入购物车数量也随之增加
- (3) 其中购买曲线与收藏曲线趋势更为接近，在一定程度上说明客户对于收藏商品更为偏爱，而购物车里的一部分商品可能只是作为一种比较或者是排除购买的作用，可以关注客户购物车和收藏的商品进行精准推送。

## ⑥ 跳失率、复购率

跳失率=只有浏览行为的用户数/总用户数

```
SELECT SUM(IF(t.fav_num=0 AND t.cart_num=0 AND t.buy_num=0,1,0))/COUNT(t.uid)*100 AS '跳失率'
FROM (
    SELECT uid,
        SUM(IF(b_type='收藏',1,0)) fav_num,
        SUM(IF(b_type='加购',1,0)) cart_num,
        SUM(IF(b_type='购买',1,0)) buy_num
    FROM user_behavior GROUP BY uid ) t;
```



跳失率
18.7330

可得跳失率约为 18.733%，说明详情页对用户吸引力仍有待提高

**复购率 = 购买次数两次及其以上的人数/有购买行为的用户总数**

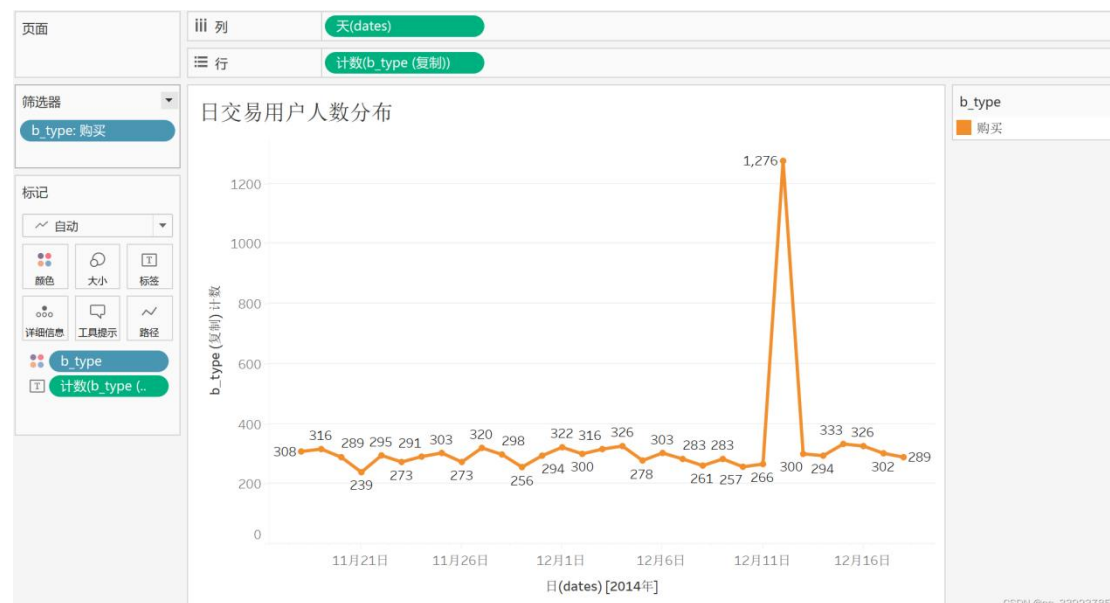
```
SELECT * FROM user_behavioror;
SELECT SUM(CASE WHEN buy_amount>1 THEN 1 ELSE 0 END) AS '复购人数',
       COUNT(uid) AS '已购人数', SUM(CASE WHEN buy_amount>1 THEN 1 ELSE 0 END)/COUNT(uid) AS '
复购率'
FROM (SELECT uid,COUNT(b_type) AS buy_amount FROM user_behavioror WHERE b_type='购买' GROUP BY
uid) AS a;
```

复购人数	已购人数	复购率
2295	4330	0.5300

复购率在 53%，用户粘性比较不错，用户平台消费依然需要加强宣传、活动优惠等，提升用户消费兴趣，提高用户粘性。

### 3、用户购买行为分析

日交易用户人数分布



可以看到，在双十二当天的交易数在 1276，是平常数量的 4 倍左右。其他日期的购买量基本处于平稳水平，这与消费者整体购买水平和消费者购买习惯等有着重要关系。

### 3、建立用户各行为转化漏斗模型

#### ①创建用户行为视图

```
CREATE VIEW userbehavior AS SELECT uid,
    COUNT(b_type) AS '用户行为总数',
    SUM(CASE WHEN b_type = '点击' THEN 1 ELSE 0 END) AS '点击量',
    SUM(CASE WHEN b_type = '购买' THEN 1 ELSE 0 END) AS '购买量',
    SUM(CASE WHEN b_type = '收藏' THEN 1 ELSE 0 END) AS '收藏数',
    SUM(CASE WHEN b_type = '加购' THEN 1 ELSE 0 END) AS '加购数'
FROM user_behavior GROUP BY uid;

SELECT * FROM userbehavior; ##计算用户各行为发生总数

SELECT SUM(点击量) AS total_pv, SUM(购买量) AS total_buy, SUM(收藏数) AS total_fav, SUM(加购数) AS
total_cart FROM userbehavior;
```

uid	用户行为总数	点击量	购买量	收藏数	加购数
98047837	174	165	0	5	4
97726136	71	69	0	0	2
98607707	153	149	2	1	1
98662432	6	6	0	0	0
98145908	10	10	0	0	0
93784494	232	223	1	6	2
94832743	11	11	0	0	0
95290487	85	80	2	3	0
96610296	61	53	2	6	0
100684618	96	90	3	2	1
100509623	281	261	4	16	0
101781721	42	39	0	3	0
103891828	201	185	6	10	0
101260672	21	20	1	0	0
104811265	18	16	1	0	1
103802946	115	108	2	1	4
104221274	369	360	2	1	6
110418606	6	6	0	0	0
116730636	147	132	4	4	7
106230218	16	12	3	0	1
103582986	178	168	0	10	0
106557109	222	209	1	11	1
110253038	46	45	0	1	0
116678892	99	98	1	0	0
103898055	38	36	2	0	0
107874169	228	204	6	16	2
11087331	18	18	0	0	0
112296733	78	73	0	5	0
132252054	22	20	1	1	0
133088851	35	35	0	0	0

(共 8477 行，这里是部分数据)

total_pv	total_buy	total_fav	total_cart
987911	10240	29427	20997

## ②计算各行为的转化率

##计算各行为的转化率

```
SELECT CONCAT(ROUND(SUM(点击量)/SUM(用户行为总数)*100,2),'%') AS total_to_pv,  
CONCAT(ROUND((SUM(收藏数)+SUM(加购数))/SUM(点击量)*100,2),'%') AS pv_to_favCart,  
CONCAT(ROUND(SUM(购买量)/SUM(点击量)*100,2),'%') AS pv_to_buy  
FROM userbehavior;
```

total_to_pv	pv_to_favCart	pv_to_buy
94.21%	5.10%	1.04%

从以上结果说明：

- 1、用户从浏览到点击的转化率只有 94.21%
- 2、用户从点击后收藏且加入购物车的转化率有 5.10%
- 3、用户从收藏及加购到再购买的转化率有 1.04%

进一步分析：

1、从用户浏览到收藏加购的转化率大于 5%，处于正常的水平，消费者关注的元素能够比较快速的让消费者看到，产生环节转化的可能。收藏加购行为到购买的转化率直接提升了 3 倍，接近 20%。这说明在收藏加购环节，消费者在对商品、服务等消费因素都比较满意，对购买的满意度逐渐提升，为最终的购买行为产生奠定了条件。因此，我们可以多从产品的交互界面、营销机制等方面提高用户收藏、加购的行为，进而提升交易量。

2、用户的流失率非常低，只有不到 6%。消费者在浏览完页面后不采取任何行为的数量占比非常小，这说明页面信息对消费者的兴趣吸引和推送服务内容上做的还是不错，在第一环节尽力留住了消费者。消费者在第一环节得以保留，为下面的环节加购、收藏、购买的行为产生提供了条件。





在进行漏斗转化图形绘制时，没有按照 sql 语句执行来绘制，而是在 Tableau 中重新创建字段进行漏斗分析，因此在阶段客户数量上有差异。如图。

#### 4、用户留存分析

**n 日后留存率=(注册后的 n 日后还登录的用户数)/第一天新增总用户数**

##用户留存分析

--第一天新用户数

```
CREATE TABLE lc AS SELECT COUNT(DISTINCT uid) AS 第一天新用户数 FROM user_behavior WHERE dates = '2014-11-18';
```

-- 第二天留存用户数

```
ALTER TABLE lc ADD COLUMN 第二天留存用户 INT;
```

```
UPDATE lc SET 第二天留存用户= (SELECT COUNT(DISTINCT uid)
```

```
FROM user_behavior WHERE dates = '2014-11-19' AND uid IN (SELECT uid FROM user_behavior WHERE dates = '2014-11-18'));
```

-- 第三天留存用户数

```
ALTER TABLE lc ADD COLUMN 第三天留存用户 INT;
```

```
UPDATE lc SET 第三天留存用户= (SELECT COUNT(DISTINCT uid)
```

```
FROM user_behavior WHERE dates = '2014-11-20' AND uid IN (SELECT uid FROM user_behavior WHERE dates = '2014-11-18'));
```

-- 第五天留存用户数

```
ALTER TABLE lc ADD COLUMN 第五天留存用户 INT;
```

```
UPDATE lc SET 第五天留存用户= (SELECT COUNT(DISTINCT uid)
```

```
FROM user_behavior WHERE dates = '2014-11-22' AND uid IN (SELECT uid FROM user_behavior WHERE dates = '2014-11-18'));
```

--第七天留存

```
ALTER TABLE lc ADD COLUMN 第七天留存用户 INT; UPDATE lc SET 第七天留存用户= (SELECT COUNT(DISTINCT
```

```
uid) FROM user_behavior WHERE dates = '2014-11-23' AND uid IN (SELECT uid FROM user_behavior WHERE dates = '2014-11-18'));
```

```
SELECT CONCAT(ROUND(100*第二天留存用户/第一天新用户数,2),'%')AS 次日留存率,
```

```
CONCAT(ROUND(100*第三天留存用户/第一天新用户数,2),'%')AS 三日留存率,
```

```
CONCAT(ROUND(100*第五天留存用户/第一天新用户数,2),'%')AS 五日留存率,
```

```
CONCAT(ROUND(100*第七天留存用户/第一天新用户数,2),'%')AS 七日留存率
```

```
FROM lc;
```

次日留存率	三日留存率	五日留存率	七日留存率
70.77%	68.60%	63.65%	65.26%

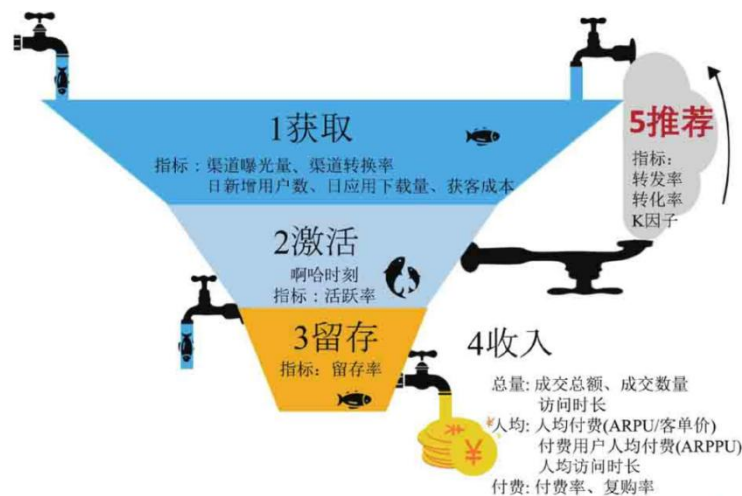
可以看出：

次日留存达 8 成左右，说明在顾客的吸引力上较强，能够抓住消费者的兴趣点，。三日留存、五日留存以及七日留存都在 6 成以上，相比次日下降幅度均不算太大。这说明，平台消费者在消费上对平台有着一定的信任度，顾客粘性较强，用户活跃度较高，平台整体处于成熟期（即上升时期）。

此外，可以再进行周留存比较，不再举例。

## 四、总结

本项目运用 MySQL+Tableau，在 1048575 条用户行为数据中进行分析，下面结合 AARRR 模型，对应用户生命周期中的 5 个重要环节，对用户行为进行总结。



### 1、用户获取

由 PV、UV 每时变化趋势图可知 18 点到 23 点是用户频繁访问的时间，也是获取更多潜在客户的黄金时间。可以用最小的获客成本进行小游戏邀请、抖音/快手社交或小视频平台合作推广、app 卖家推送等增加渠道曝光量进而增加渠道转化率。本项目中跳失率约为 18.733%，说明详情页对用户吸引力仍有待提高，可以进行精准人群推广、视觉优化等方式降低跳失率。

### 2、用户激活

本项目中 2014 年 11 月 18 日-2014 年 12 月 18 日 DAU 较为平稳，12 月 11 日-12 日由于“双十二”购物节活动增长迅速，也符合活动预期。本项目分析了用户的四种行为，包括点击、购买、加入购物车、收藏。点击量占总行为的 94.21%，而收藏和加入购物车的用户仅占 5.10%，最终购买的用户只有 1.04%，转折点在加入购物车和收藏，要考虑为什么用户浏览了那么多商品却只有 5.10%的用户加入购物车或收藏，可能这中间用户花了太多的时间却没有挑选到满意的商品，那么有一些客户就流失掉了。可以采取问卷调查的方式分析用户行为推送用户可能感兴趣的物品、优化购物界面使界面布局更加人性化等方式更好的满足用户需求，激活更多用户。

### 3、提高留存

在本项目中用户留存率相对稳定，七日留存率略有回升，可能是因为当天 11 月 23 日是星期五，正好是客户活跃度上升的一个转折点，可能是有购物优惠活动。

为了提高留存，可以让用户养成习惯，在一定程度上增加用户对产品的依赖性。①每日给用户推送可能感兴趣的商品；②一些签到的小游戏或者签到领积分活动，增加用户登录的频率，进而增加用户浏览时长；③对于年购买量比较大的用户，为了增加这些用户的粘性，可以推出 vip 服务，让这些用户享受折扣和优惠券，进而提高高价值用户的留存率和对平台的忠诚度；④大力推广直播跟用户互动，计算亲密指数并进行排行，可以让用户与店主之间建立信任，进而增加客户的忠诚度。

#### 4、增加收入

为了增加收入要提高成交转化率、复购率以及加购或收藏产品的购买率。①提高成交转化率，优化用户推荐系统，或者邀请明星做测评推广；②通过分析找出价值用户的购买偏好，产品和类目等，给价值用户制定个性化的产品推荐，从而提高用户体验，进而提高购买率。

#### 5、用户传播

要想让用户自发传播，那这个产品口碑一定要非常好，而且能够充分满足用户的需求，甚至给用户带来期望的需求。所以，要好好打磨产品。可以采用一些增长策略，像现在比较普遍的老带新裂变营销，把握用户心理策划一些噱头式的活动等。还可以拼团、分享得优惠券、分享到朋友圈额外送积分等方式，鼓励用户进行分享和传播。

## 续用户淘宝行为数据分析

多年以来电子商务业务快速发展，尤其是移动客户端发展迅猛，移动互联网时代的到来让原本就方便快捷的网上购物变得更加便利，而淘宝作为国内最大的电商交易平台更是具有巨大的流量优势。淘宝 APP 的功能日益复杂，但都离不开最基本的收藏、购物车和购买功能，本文利用 sql、Tableau 对淘宝用户行为数据进行分析，通过用户行为分析业务问题。

### 一、提出问题

前述文章内容“用户行为分析”已经解答了 AARRR 漏斗模型分析的问题以及不同时段用户的行为数据分析。这里再做两个问题延展，作为本数据集运用分析的结尾。

分别为：

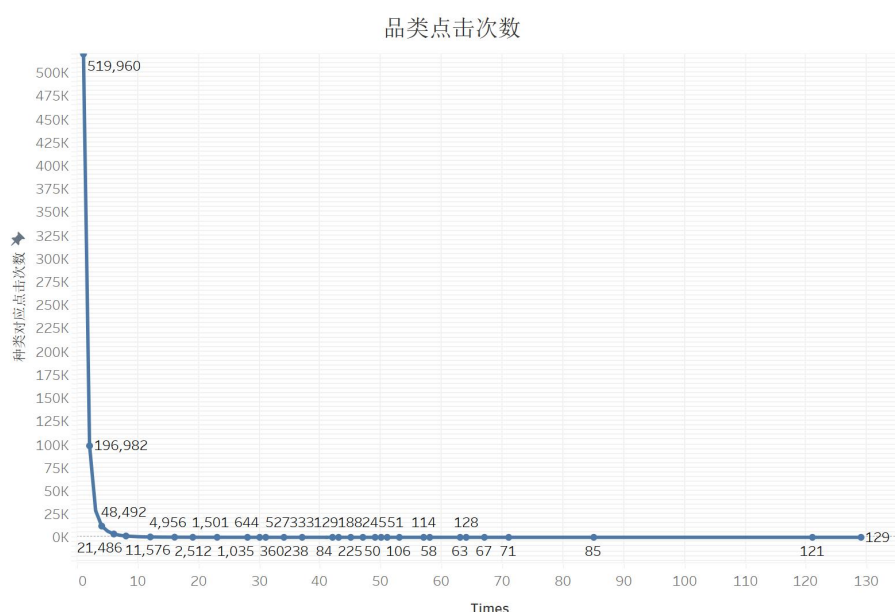
- 1.找到用户对商品的偏好，找到针对不同商品的营销策略
- 2.找出最具价值的核心付费用户群，对这部分用户的行为进行分析（RFM 模型分析）

问题：

- 1.找到用户对商品的偏好，找到针对不同商品的营销策略

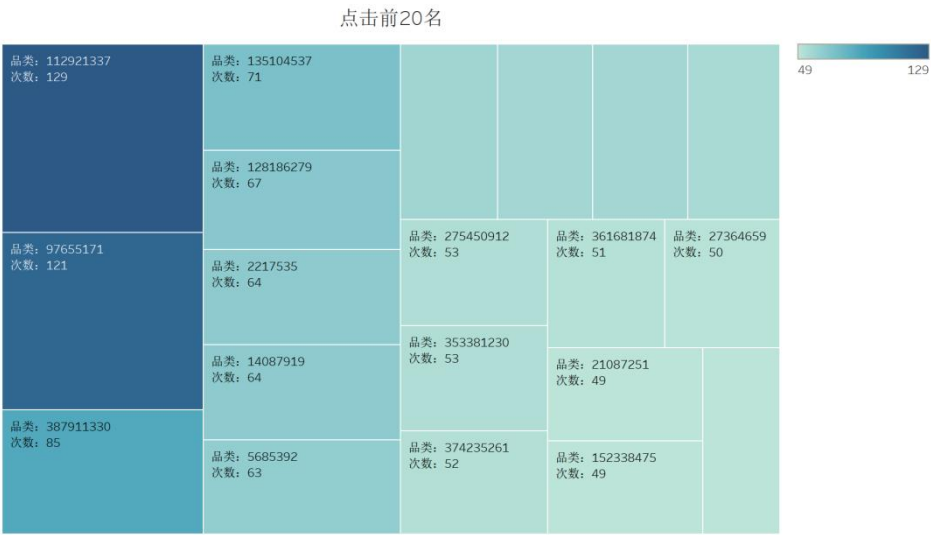
统计所有商品的购买次数，同时找到购买次数、浏览次数、收藏次数和加入购物车次数最多的商品。

```
SELECT item_id,COUNT(uid) AS 'times' FROM user_behavior WHERE b_type='点击' GROUP BY item_id;
SELECT item_id,COUNT(uid) AS 'times' FROM user_behavior WHERE b_type='收藏' GROUP BY item_id;
SELECT item_id,COUNT(uid) AS 'times' FROM user_behavior WHERE b_type='加购' GROUP BY item_id;
SELECT item_id,COUNT(uid) AS 'times' FROM user_behavior WHERE b_type='购买' GROUP BY item_id;
```

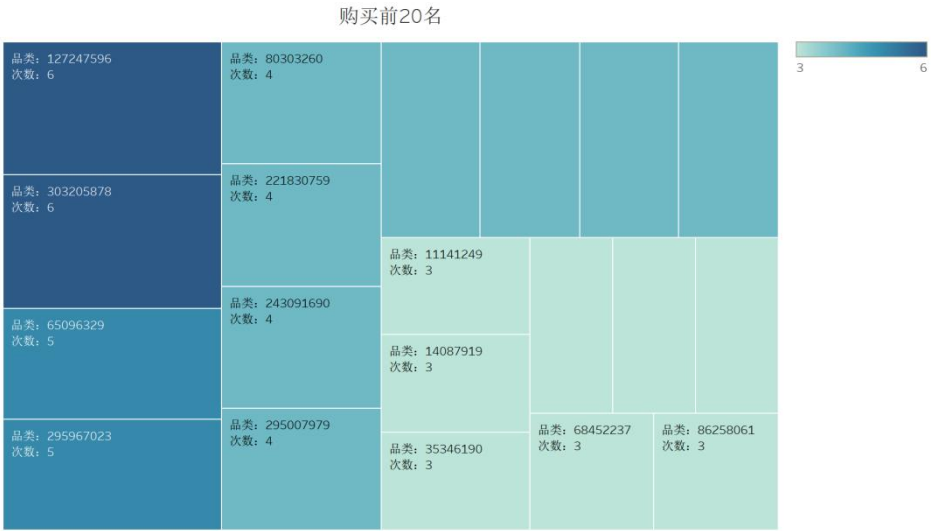




在本次统计的数据中，共有 701611 种。其中，只点击一次的产品 519960 种，占到 76.79%；点击了两次的产品有 98491 种，占到 14.55%。这说明商品浏览量的增长主要依靠长尾商品的累积效应，而非爆款商品的带动。

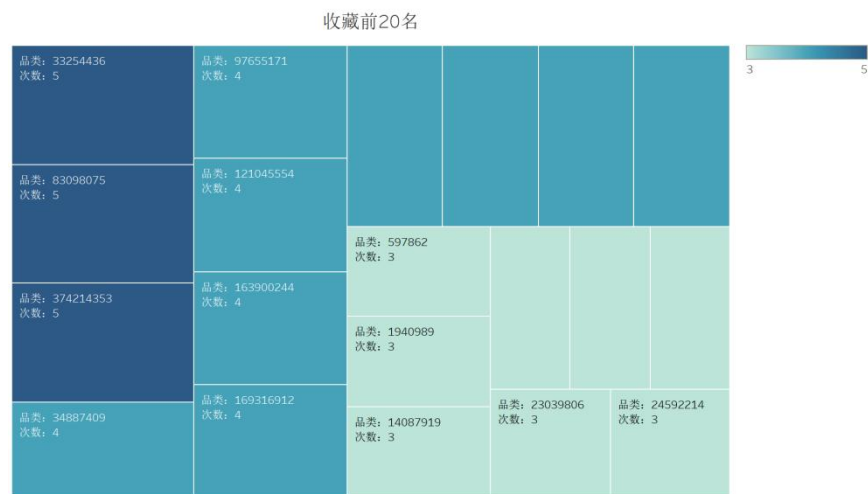


点击前 20 名，品类 112921337 点击了 129 次，97655171 点击了 121 次，其余均在 100 次以下。

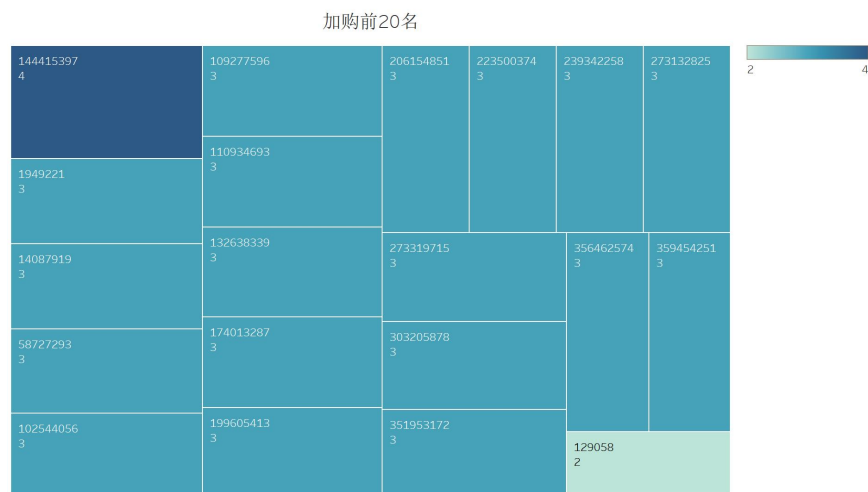




列出销量前 20 位的商品，item\_ID 为 127247596、303205878 的商品购买次数最高，为 6 次。



我们看到收藏数最高的商品为 33254436、83098075、374214353，收藏数为 5。进一步分析，销量最高的商品没有进入收藏前 20 名，说明这些吸引用户更多注意力的商品没有很好的转化为实际销量。



在加入购物车次数最多的前 20 个商品中，可以看到加购第一而不是销量第一，说明加购数与销量的关系并不是很明显直接。（这里截取的数据 104 万条，截取时段为一个月，可能存在数据量不足或者截取不够的原因，加大数据量可能关系会更明显）

## 2.基于 RFM 理论找出有价值的用户

### 2.1 我们首先统计出用户购买商品的最大值和最小值

数据源中没有相关的金额数据，因此通过 R 和 F 的数据对客户价值进行打分。

R-最近购买时间

用户数据的时间范围为一个月，最近购买时间的区间为 0-30，将其分为 5 档，0-6,7-12,13-18,19-24,25-30 分别对应评分 0 到 4；

## F-消费频率

付费用户中消费次数从低到高为 1-15 次，将其分为 5 档，1-3，4-6，7-9，10-12,13-15 分别对应评分 0 到 4

```
CREATE VIEW pay_E AS
SELECT uid,DATEDIFF('2014-12-18',MAX(dates)) AS B , COUNT(*) AS A
FROM user_behavinor WHERE b_type='购买' GROUP BY uid;
###创建视图
SELECT uid,
(CASE WHEN B BETWEEN 25 AND 30 THEN 0 WHEN B BETWEEN 19 AND 24 THEN 1 WHEN B BETWEEN 13 AND 18 THEN 2 WHEN B BETWEEN 7 AND 12 THEN 3 WHEN B BETWEEN 0 AND 6 THEN 4 ELSE NULL END) AS R,
(CASE WHEN A BETWEEN 1 AND 3 THEN 0 WHEN A BETWEEN 4 AND 6 THEN 1 WHEN A BETWEEN 7 AND 9 THEN 2 WHEN A BETWEEN 10 AND 12 THEN 3 WHEN A BETWEEN 13 AND 15 THEN 4 ELSE NULL END) AS F
FROM pay_E ORDER BY R DESC,F DESC;
```

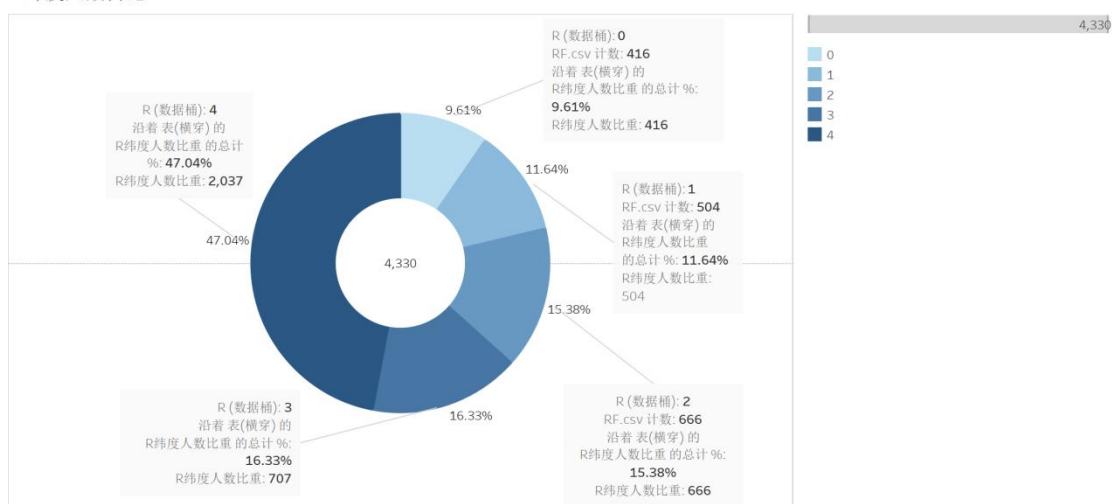
将 R 和 F 的评分合并后结果：

uid	R	F
110976685	4	4
102825286	4	4
35389479	4	4
71714329	4	4
1926899	4	4
20379505	4	4
4120403	4	4
50244004	4	3
85336332	4	3
4361577	4	3
138735272	4	3
110883373	4	3
109613455	4	3
33797179	4	3
79253885	4	3
14281644	4	3
22393155	4	3
17821367	4	3

CSDN @qq\_33923785

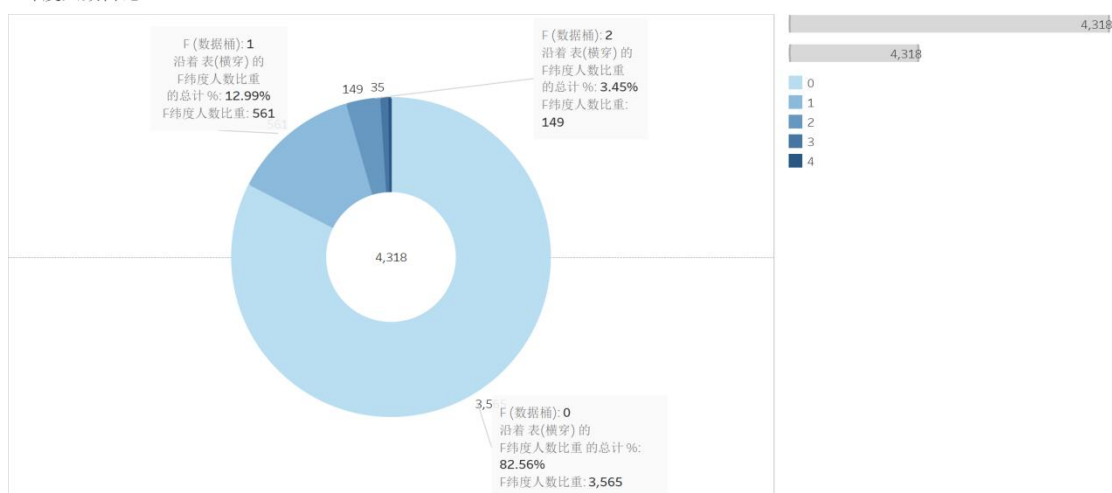
（共 4430 行，列出部门）

R纬度人数占比



从R时间间隔饼图可以看出，间隔评分在4分的占到了47.04%，共计2037人，趋近于一般人数。说明消费者在消费间隔上较小，消费频率集中。从侧面也看出，消费者对消费的热度很大，兴趣很高。除4我i，评分3的人数达到707人，约占16.33%，评分2的人数有666人，约占15.38%。2分和3分人数共有约31.71%，这部分消费群体时间间隔在半个月左右，是处于中等消费频率的区间。针对这部分人群，应该细化市场，将消费者消费兴趣提高，比如可以做一些活动游戏，或者消费优惠活动，让消费者可以累积消费，提高热度。

F纬度人数占比



F纬度评分，位于评分0的人数最多，占到了82.56%。说明购买次数在1-3次的人数是最多的。考虑到消费水平的约束以及消费习惯的固定，这部分消费分析是较为合理的。评分1的人数在561，占到12.99%，位于第二位。这部分人群消费频率在4-6次，处于中等水平。这部分人群是应该加以重要挖掘的客户，应该重点加深精准活动推送，优惠券发放，节假日活动提醒等，提高消费者消费频率。

## 2.2 R、F 维度综合评分进行用户价值分层

###R、F 综合打分

# RF 综合打分

```
CREATE VIEW dafen AS
SELECT uid,
(CASE WHEN B BETWEEN 25 AND 30 THEN 0
      WHEN B BETWEEN 19 AND 24 THEN 1
      WHEN B BETWEEN 13 AND 18 THEN 2
      WHEN B BETWEEN 7 AND 12 THEN 3
      WHEN B BETWEEN 0 AND 6 THEN 4 ELSE NULL END) AS R,
(CASE WHEN A BETWEEN 1 AND 3 THEN 0
      WHEN A BETWEEN 4 AND 6 THEN 1
      WHEN A BETWEEN 7 AND 9 THEN 2
      WHEN A BETWEEN 10 AND 12 THEN 3
      WHEN A BETWEEN 13 AND 15 THEN 4 ELSE NULL END) AS F
FROM pay_E ORDER BY R DESC, F DESC;
```

##1. 先创建打分视图

##2. 创建综合视图

```
CREATE VIEW rf_score AS SELECT uid AS '用户ID', R AS 'R 维度评分', F AS 'F 维度评分', (R + F) AS 'RF
维度综合评分'
```

FROM dafen;

# RF 维度综合用户评分

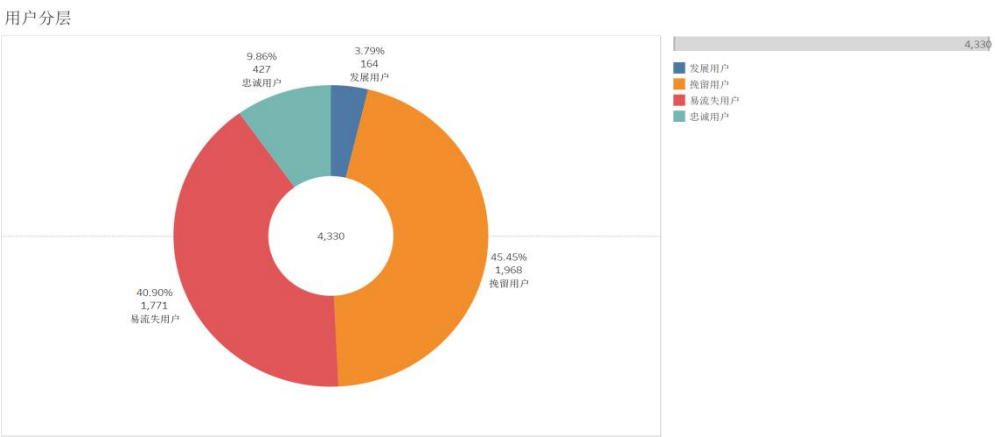
```
SELECT (CASE WHEN RF 维度综合评分 BETWEEN 1 AND 3 THEN '易流失用户'
      WHEN RF 维度综合评分 BETWEEN 4 AND 5 THEN '挽留用户'
      WHEN RF 维度综合评分 BETWEEN 6 AND 8 THEN '发展用户' ELSE '忠诚用户' END) '用户分层', COUNT(*) '
用户数量'
```

FROM rf\_score GROUP BY 用户分层;

结果如下:

<input type="checkbox"/>	用户分层	用户数量
<input type="checkbox"/>	挽留用户	1968
<input type="checkbox"/>	易流失用户	1771
<input type="checkbox"/>	忠诚用户	427
<input type="checkbox"/>	发展用户	161

CSDN@qq\_33923785



分析：

1、将 R 维度评分+F 维度评分=RF 综合评分的指标，

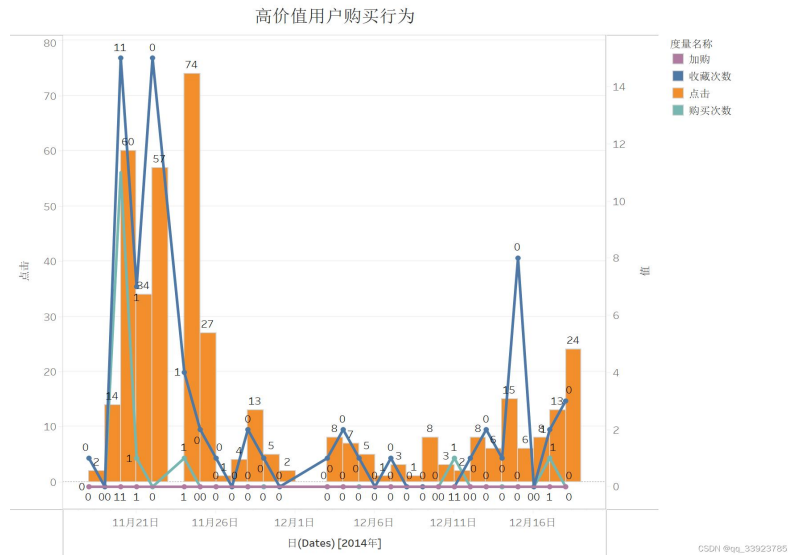
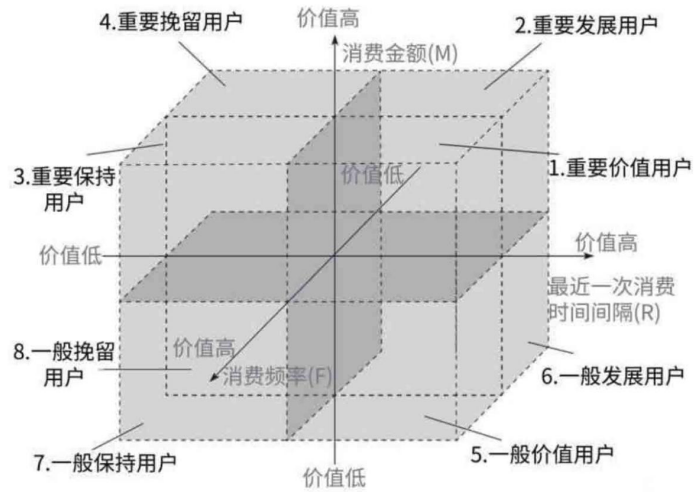
客户分类	评分
易流失用户	1-3分
挽留用户	4-5分
发展用户	6-7分
忠诚用户	7分以上

如图，可以看出，大部分用户集中在易流失用户和挽留用户上，共计占到了 86.35%。  
这里，借鉴数据分析猴子一书中的 RFM 模型分析措施：

客户分类	表现	措施
易流失用户	消费时间距离近，消费频率低	提高消费频率、发放短期小额优惠券
挽留用户	消费时间距离远，消费频率低	联系用户，调查挽回
发展用户	消费时间距离远，消费频率高	邮箱推送,app提醒，促销
忠诚用户	消费时间距离近，消费频率高	提高vip服务

最后，借鉴数据分析猴子一书中的 RFM 模型解释图：我们得知用户的 RFM 分数，用户 ID110976685、102825286、35389479、71714329、1926899、20379505、4120403 的 R、F 均是 4 分，是体系中的最价值用户，需要重点关注，并且活动投放时需谨慎对待，不要引起用户反感。对于 R 值为 4 而 F 值为 0 的用户，用户粘性不强而消费时间间隔较短，运营活动可以重点针对这部分用户，提高用户使用产品的频率，可以通过拼团打折、积分兑换等活动唤起用户注意力。





我们提取出评分最高的用户 110976685 行为数据,发现该用户几乎每天都产生了点击和收藏行为,而购物车的使用频率非常低,为 0 次。考虑到这是双十一刚刚过去一周左右,用户的购买热情与商家的优惠活动接近结束末期,用户的购物出现短暂的高涨期,随后会逐渐趋于正常。对此,可以根据该用户的活跃规律进行相关活动的推送,由于商品种类数据为脱敏数据,此处无法分析该用户购买的商品类型。

## 二、结论与建议

本文分析了淘宝 APP 用户行为数据共 104 万条，从两个角度提出业务问题，和 RFM 模型分析数据给出如下结论和建议。

### 1. 找到用户对不同种类商品的偏好，找到针对不同商品的营销策略

商品售卖主要依靠长尾商品的累积效应，而非爆款商品的带动。销量最高的商品浏览数都没排进前 20 位，而浏览次数最高的商品甚至没有进入销量前 20，说明这些吸引用户更多注意力的商品没有很好的转化为实际销量。

针对浏览量高而销量不高的这部分商品，需要提高的是用户从点击进入商品详情页到最终购买的体验。作为商家端可以从以下几个方面提高销售额：

（1）商品详情页的实际价格是否相比展示价格偏差过大，有的商家为了吸引用户点击在商品展示页投放的价格具有较强吸引力，但实际价格偏高，在用户心中反而引起反感

（2）详情页的信息流展示是否合理，是否将用户最想看到的部分置于容易看到的位置，便于信息的获取

（3）优化商品展示的形式，利用视频等方式给用户更直观的感受，提高照片的美观程度

（4）评论区评价管理，尤其对于差评区的用户反馈进行认真对待，提高自身服务质量

### 2. 通过 RFM 模型找出最具价值的核心付费用户群，对这部分用户的行为进行分析

R 和 F 评分都很高的用户是体系中的最有价值用户，需要重点关注，并且活动投放时需谨慎对待，不要引起用户反感。对于 R 值为 4 而 F 值为 0 的用户，用户粘性不强而消费时间间隔较短，运营活动可以重点针对这部分用户，提高用户使用产品的频率，可以通过拼团打折、积分兑换等活动唤起用户注意力。