淘宝用户行为分析(SQL)

#1.建库建表

```
create database taobao;
use taobao;
drop table if exists tb;
```

数据来源于阿里天池，数据集包含了淘宝 app2017 年 11 月 25-12 月 3 日的用户行为数据，包括浏览、收藏、加购和购买。每一条数据代表用户的行为记录，包含用户编号 user_id,商品编号 item_id，商品类型 category_id，行为类型 behavior_type 以及时间戳 Time_stamp。由于数据量上亿，文件太大，因此从中仅选用了 100 万样本。
行为类型包括四种：pv（浏览）、cart（加购）、fav（收藏）、buy（购买）

```
create table tb(
user_id varchar(255)   not null,
item_id varchar(255),
category_id varchar(255),
behavior_type varchar(10),
Time_stamp int(11)
);

#将 csv 文件导入数据库
load data infile 'F:/TianChi/OutPut/UserBehavior-009.csv'
into table tb
fields terminated by ','
optionally enclosed by '"'
escaped by '"'
lines terminated by '\r\n';

#复制一张表,以免丢失
create table ub select * from tb;
select * from ub;

SHOW DATABASES;
show tables;
select count(*) from tb;
```

#2.查看是否有缺失值

```
select * from tb
where user_id is null or item_id is null or category_id is null
or behavior_type is null or Time_stamp is null;#结果无缺失值
```

	user_id	item_id	category_id	behavior_type	Time_stamp	dates	times	hours

#另一种方法
select count(1) from tb where user_id is null;

#查看是否有重复值
select count(*) from tb
group by user_id,item_id,category_id,behavior_type,time_stamp
having count(*) >1;

Empty set (7.05 sec)

#3.时间戳的处理
SET SQL_SAFE_UPDATES = 0;
#(1)添加 dates 列，记录行为发生日期
alter table tb add dates varchar(255);
update tb set dates=from_unixtime(Time_stamp,'%Y-%m-%d');

#(2)添加时间列，记录发生时间
alter table tb add times varchar(255);
update tb set times =from_unixtime(Time_stamp,'%H:%i:%s');

#4.异常值处理
#(1)查看行为类型是否存在异常
select behavior_type from tb
where behavior_type not in ('pv','cart','buy','fav');

#(2)查看 Timestamp 是否存在异常
select dates from tb where dates not between '2017-11-25' and '2017-12-03';
#删除错误数据
delete from tb where dates not between '2017-11-25' and '2017-12-03';

#检查数据概况
select count(distinct user_id) as '用户总数',
count(distinct item_id) '商品总数',
count(distinct category_id) '商品类型总数',
count(distinct behavior_type) '行为类型总数',
count(distinct dates) as '天数'
from tb;

```
+----------+---------+-------------+-------------+------+
| 用户总数  | 商品总数 | 商品类型总数 | 行为类型总数 | 天数  |
```

```
+----------+---------+-----------+------------+------+
|      9739  |    398972 |          5793 |             4 |     9  |
+----------+----------+-------------+-------------+------+
```

#数据清洗完毕以后，进行分析和可视化

#5.分析

#5.1 分析用户每日行为规律

#每日浏览量,用户数

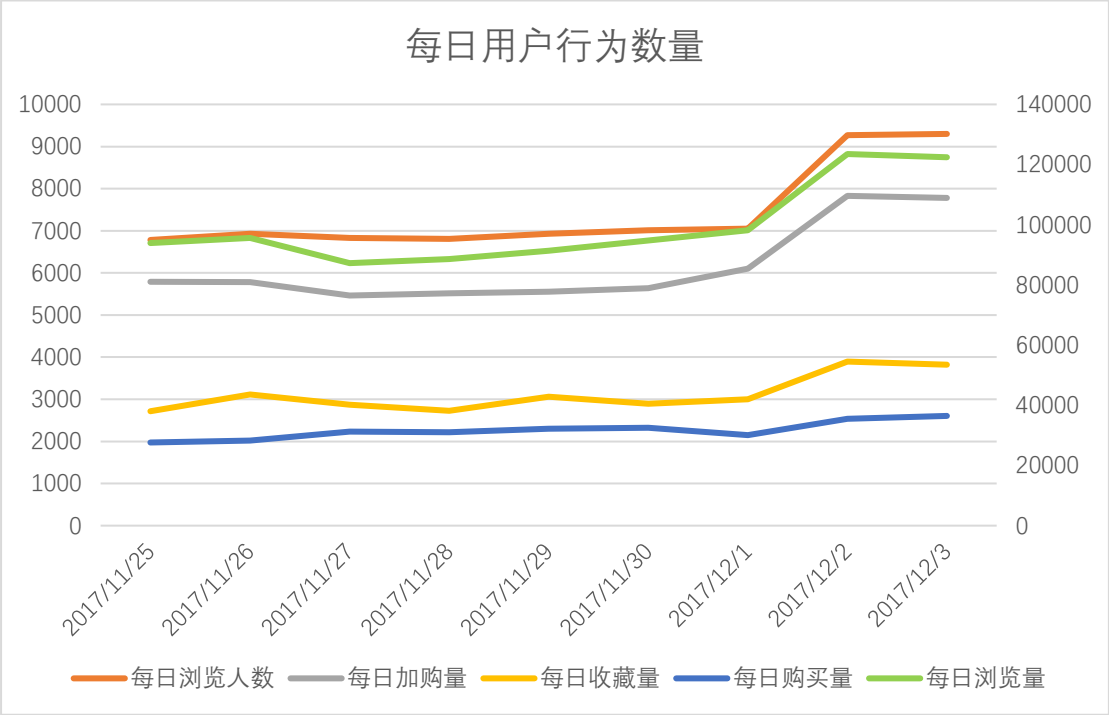
```
select dates,count(behavior_type) as '每日浏览量',count(distinct user_id) as '每日浏览人数'
from tb
where behaviortype='pv'
group by dates
order by dates;
```

```
+------------+------------+-------------+
| dates      | 每日浏览量 | 每日浏览人数 |
+------------+------------+-------------+
| 2017-11-25 |      93931 |        6782 |
| 2017-11-26 |      95657 |        6928 |
| 2017-11-27 |      87243 |        6828 |
| 2017-11-28 |      88638 |        6810 |
| 2017-11-29 |      91334 |        6930 |
| 2017-11-30 |      94735 |        7010 |
| 2017-12-01 |      98138 |        7054 |
| 2017-12-02 |     123514 |        9271 |
| 2017-12-03 |     122446 |        9300 |
+------------+------------+-------------+
9 rows in set (4 min 39.32 sec)
```

#每日加入购物车数量

```
+------------+------------+
| dates      | 每日加购量 |
+------------+------------+
| 2017-11-25 |       5786 |
| 2017-11-26 |       5784 |
| 2017-11-27 |       5463 |
| 2017-11-28 |       5516 |
| 2017-11-29 |       5550 |
| 2017-11-30 |       5638 |
| 2017-12-01 |       6102 |
| 2017-12-02 |       7829 |
| 2017-12-03 |       7779 |
+------------+------------+
```

```
+------------+------------+
| dates      | 每日收藏量 |
+------------+------------+
| 2017-11-25 |       2716 |
| 2017-11-26 |       3113 |
| 2017-11-27 |       2873 |
| 2017-11-28 |       2726 |
| 2017-11-29 |       3057 |
| 2017-11-30 |       2891 |
| 2017-12-01 |       2999 |
| 2017-12-02 |       3895 |
| 2017-12-03 |       3818 |
+------------+------------+
+------------+------------+
| dates      | 每日购买量 |
+------------+------------+
| 2017-11-25 |       1974 |
| 2017-11-26 |       2022 |
| 2017-11-27 |       2229 |
| 2017-11-28 |       2220 |
| 2017-11-29 |       2299 |
| 2017-11-30 |       2323 |
| 2017-12-01 |       2151 |
| 2017-12-02 |       2536 |
| 2017-12-03 |       2605 |
+------------+------------+
9 rows in set (3.01 sec)
```



#每日购物用户占比
select a.dates,count(distinct b.user_id)/count(distinct a.user_id) as '每日购物用户占比'
from tb a
left join (select * from tb where behavior_type='buy') b
on a.user_id=b.user_id and a.item_id=b.item_id and a.time_stamp=b.time_stamp
group by a.dates;

```
+----------+-----------------+
| dates      | 每日购物用户占比 |
+----------+-----------------+
| 2017-11-25 |          0.1888 |
| 2017-11-26 |          0.1857 |
| 2017-11-27 |          0.2014 |
| 2017-11-28 |          0.2005 |
| 2017-11-29 |          0.2054 |
| 2017-11-30 |          0.2038 |
| 2017-12-01 |          0.1950 |
| 2017-12-02 |          0.1804 |
| 2017-12-03 |          0.1840 |
+----------+-----------------+
```

#每日人均交易数
select a.dates,count(b.user_id)/count(distinct a.user_id) as '每日人均交易数'
from tb a
left join (select * from tb where behavior_type='buy') b

```
on a.user_id=b.user_id and a.item_id=b.item_id and a.time_stamp=b.time_stamp
group by a.dates;
+------------+----------------+
| dates      | 每日人均交易数 |
+----------+--------------+
| 2017-11-25 |         0.2830 |
| 2017-11-26 |         0.2837 |
| 2017-11-27 |         0.3173 |
| 2017-11-28 |         0.3157 |
| 2017-11-29 |         0.3219 |
| 2017-11-30 |         0.3207 |
| 2017-12-01 |         0.2959 |
| 2017-12-02 |         0.2651 |
| 2017-12-03 |         0.2725 |
+----------+--------------+
```

11 月 25-11 月 26 与 12 月 2 日-12 月 3 日均为周末，相比前一个周末，12 月的周末浏览量和加购量明显更多，但最终的购买量没有明显增加，推测有可能是双十二活动预热和服饰焕新所致。

注：

- 母婴冬季保暖节（11/20-11/30）
- 黑色星期五（11/22-11/26）
- 咖啡节（11/23-11/27）
- 火拼节（11/27-11/29）
- “双十二”预热阶段（12/1-12/8）
- 服饰焕新（12/1-12/6）

```
#从每日不同时段分析用户行为规律：增加一个字段'hours'
select hours,count(*) as '总浏览量' from tb
where behavior_type='pv'
group by hours order by hours;
+-------+----------+
| hours | 总浏览量 |
+-------+----------+
| 00    |    30396 |
| 01    |    13852 |
| 02    |     7984 |
| 03    |     5636 |
| 04    |     4969 |
| 05    |     6060 |
| 06    |    12071 |
| 07    |    22225 |
| 08    |    30016 |
| 09    |    37185 |
| 10    |    43649 |
```

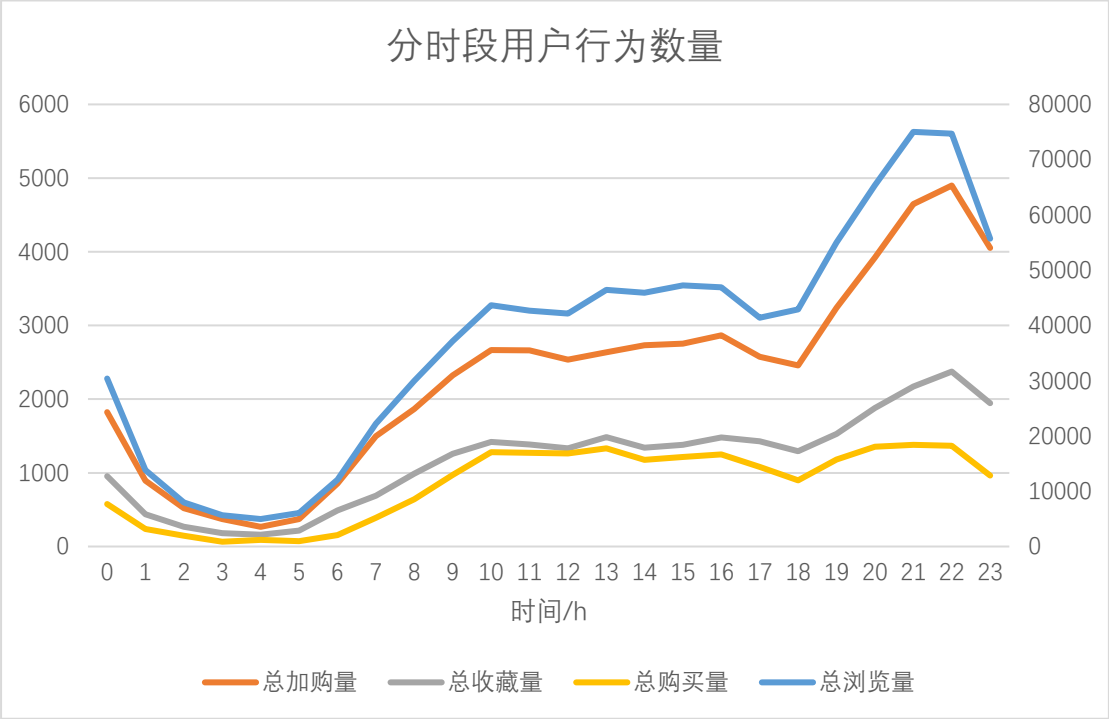
11	42704
12	42168
13	46453
14	45928
15	47250
16	46880
17	41418
18	42911
19	55026
20	65406
21	75030
22	74687
23	55732
+------+---------+

#总加购量
select hours,count(*) as '总加购量' from tb
where behavior_type='cart'
group by hours order by hours;
+-------+----------+
| hours | 总加购量 |
+-------+----------+
00	1824
01	892
02	522
03	372
04	267
05	374
06	855
07	1499
08	1869
09	2323
10	2666
11	2663
12	2538
13	2634
14	2730
15	2754
16	2867
17	2574
18	2459
19	3241
20	3922
21	4649

```
| 22      |        4899 |
| 23      |        4054 |
+-------+----------+
24 rows in set (3.08 sec)

#总收藏量
select hours,count(*) as '总收藏量' from tb
where behavior_type='fav'
group by hours order by hours;
+-------+------------+
| hours  | 总收藏量 |
+-------+------------+
| 00      |            956 |
| 01      |            438 |
| 02      |            266 |
| 03      |            180 |
| 04      |            160 |
| 05      |            217 |
| 06      |            490 |
| 07      |            689 |
| 08      |            988 |
| 09      |           1259 |
| 10      |           1418 |
| 11      |           1384 |
| 12      |           1333 |
| 13      |           1484 |
| 14      |           1342 |
| 15      |           1381 |
| 16      |           1479 |
| 17      |           1429 |
| 18      |           1294 |
| 19      |           1529 |
| 20      |           1879 |
| 21      |           2172 |
| 22      |           2374 |
| 23      |           1947 |
+-------+------------+
24 rows in set (3.04 sec)

#总购买量
select hours,count(*) as '总购买量' from tb
where behavior_type='buy'
group by hours order by hours;
```

hours	总购买量
00	576
01	236
02	148
03	65
04	88
05	73
06	155
07	388
08	643
09	970
10	1279
11	1272
12	1264
13	1334
14	1178
15	1214
16	1252
17	1081
18	898
19	1181
20	1355
21	1380
22	1367
23	962

24 rows in set (2.66 sec)



晚上 19：00-22：00 各项行为总量呈现明显上升趋势，21：00-22：00 用户行为总量达到顶峰状态，22：00 至次日凌晨 4 点，各项用户行为数量均呈现明显下降趋势，多项用户行为于次日凌晨 4 时达到最低。晚上 19 :00 过后为多数用户休息时间，从一天工作中解放出来，有时间浏览购物，凌晨为多数用户睡眠时间，因此用户行为总量降低。

#5.2 用户留存分析

对从 11 月 25 日开始到 12 月 3 期间用户的新增与留存分析。首日（11 月 25 日）的新用户为当天所有存在任意行为的用户，其余日期的新用户定义为此前未出现任何行为的用户，而当天存在任何行为的用户。

```
#找出第一次使用的用户与日期
select user_id,min(dates) as firstday from tb
group by user_id;

#用户所有使用时间
select user_id,dates from tb
group by user_id,dates;

#将用户 id、使用时间和首次使用时间放在一张虚拟表
create view v_time as
select a.user_id,a.dates,b.firstday
from (select user_id,dates from tb group by user_id,dates) a
join
(select user_id,min(dates) as firstday from tb group by user_id) b
on a.user_id=b.user_id
order by a.user_id,a.dates;
```

```
#获取第一次使用时间和后续使用时间间隔
create view v_diff as
select user_id,dates,firstday,datediff(dates,firstday) as diff_day from v_time;


#计算留存日的用户数量
create view cus_qua as
select firstday,
sum(case when diff_day=0 then 1 else 0 end) as '当日新增用户',
sum(case when diff_day=1 then 1 else 0 end) as '1 日后留存用户',
sum(case when diff_day=2 then 1 else 0 end) as '2 日后留存用户',
sum(case when diff_day=3 then 1 else 0 end) as '3 日后留存用户',
sum(case when diff_day=4 then 1 else 0 end) as '4 日后留存用户',
sum(case when diff_day=5 then 1 else 0 end) as '5 日后留存用户',
sum(case when diff_day=6 then 1 else 0 end) as '6 日后留存用户',
sum(case when diff_day=7 then 1 else 0 end) as '7 日后留存用户',
sum(case when diff_day=8 then 1 else 0 end) as '8 日后留存用户'
from v_diff
group by firstday order by firstday;

#计算留存率(round(*,3)四舍五入，保留 3 位小数)
select firstday,当日新增用户,
        concat(round(1 日后留存用户/当日新增用户,3)*100,'%') as '1 日留存率',
        concat(round(2 日后留存用户/当日新增用户,3)*100,'%') as '2 日留存率',
        concat(round(3 日后留存用户/当日新增用户,3)*100,'%') as '3 日留存率',
        concat(round(4 日后留存用户/当日新增用户,3)*100,'%') as '4 日留存率',
        concat(round(5 日后留存用户/当日新增用户,3)*100,'%') as '5 日留存率',
        concat(round(6 日后留存用户/当日新增用户,3)*100,'%') as '6 日留存率',
        concat(round(7 日后留存用户/当日新增用户,3)*100,'%') as '7 日留存率',
        concat(round(8 日后留存用户/当日新增用户,3)*100,'%') as '8 日留存率'
        from cus_qua order by firstday;
```

firstday	当日新增用户	1日留存率	2日留存率	3日留存率	4日留存率	5日留存率	6日留存率	7日留存率	8日留存率
2017-11-25	6976	79.100%	76.800%	76.500%	76.500%	77.000%	77.300%	98.500%	98.500%
2017-11-26	1610	65.200%	65.100%	66.600%	66.300%	67.200%	98.000%	97.300%	0.000%
2017-11-27	621	59.700%	63.000%	67.800%	66.700%	97.700%	97.400%	0.000%	0.000%
2017-11-28	278	58.600%	68.700%	66.500%	96.800%	98.600%	0.000%	0.000%	0.000%
2017-11-29	178	65.200%	72.500%	95.500%	97.200%	0.000%	0.000%	0.000%	0.000%
2017-11-30	74	94.600%	93.200%	97.300%	0.000%	0.000%	0.000%	0.000%	0.000%
2017-12-01	1	100.000%	100.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
2017-12-02	1	100.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%

由于没有前置的用户数据，因此得到的新增用户与实际略有差异，但观察新增数据发现，每日新增用户呈现下降趋势，并且留存率在 58%以上，12 月 2 日和 3 日的留存率最高，符合双十二预热带来的用户关注度。

```
#用户购买情况
```

#这里主要从复购率角度，复购率指的是产生两次或者两次以上的购买的用户占购买用户的占比
#每位用户购买次数
```
create view rebuy as
select    user_id,count(behavior_type) as buy_times
from tb
where behavior_type='buy'
group by user_id
having count(behavior_type)>=2
order by buy_times desc;
```

#计算复购率
```
select count(distinct user_id) from rebuy;#计算购买次数两次及以上的用户数量

select count(distinct user_id) from tb
where behavior_type='buy';#所有购买行为的用户数

select (select count(distinct user_id) from rebuy)/(select count(distinct user_id)from tb
where    behavior_type='buy') as '复购率';
```

复购率
0.6621

#复购用户中购买人次，人数
```
select buy_times as '购买人次',count(distinct user_id) as '购买人数'
from rebuy
group by  购买人次
order by  购买人次  desc;
```

购买人次	购买人数
2	1574
3	1041
4	613
5	384
6	259
7	168
8	100
9	73
10	46
11	34

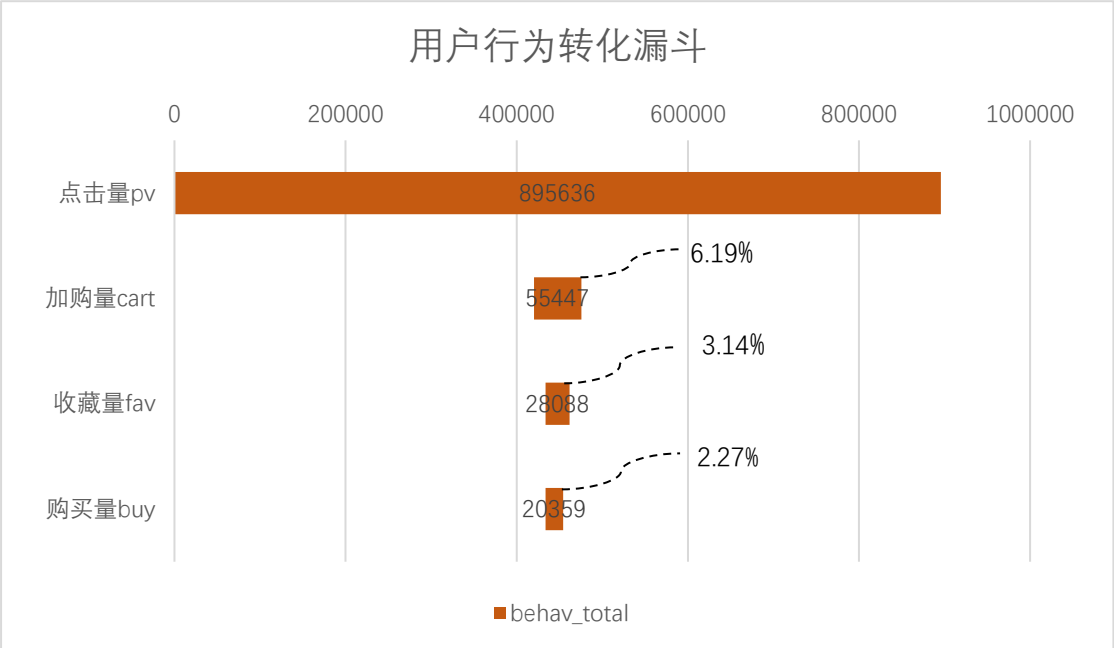
购买人次	购买人数
29	1
30	1
31	2
32	1
36	1
43	1
57	1
65	1
69	1
72	1

从上面可以看到，11/25-12/3 大部分复购用户的购买次数集中在 2-6 次，有 9 位用户在这 9 天的购买次数达到 30 次以上，最高达到了 72 次；说明淘宝用户忠诚度较高。

#5.3 用户行为路径分析
#不同行为转化率
```
select behavior_type,count(behavior_type) as behav_total
```

from tb
group by behavior_type;

behavior_type	behav_total
pv	895636
fav	28088
buy	20359
cart	55447



从转化漏斗可以看出，用户从点击到加购的转化率为 6.19%，从点击到收藏的转化率为 3.14%，从点击到购买的转化率为 2.27%。

#转化率低的原因？

- 假设 1：用户只是浏览商品而没有购买商品？
- 假设 2：用户加购/收藏到购买的转化率更高？

#假设 1：用户只是浏览商品而没有购买商品
#访客行为转化率
select behavior_type,count(distinct user_id) '访客数量'
from tb group by behavior_type;

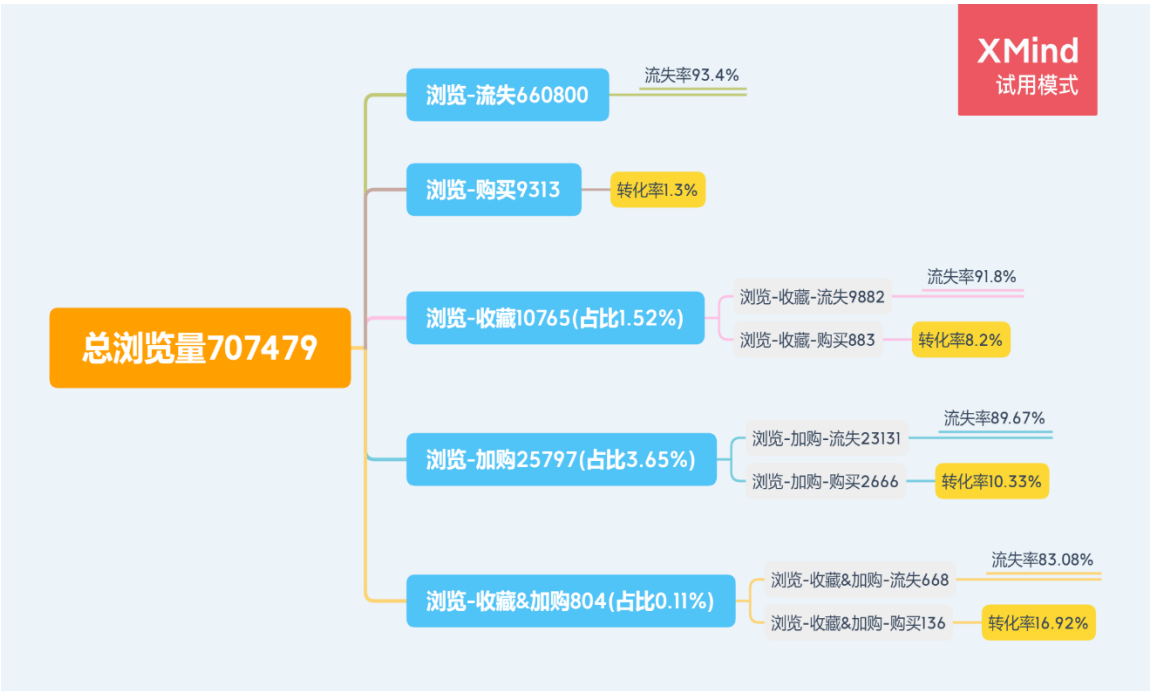
behavior_type	访客数量
fav	3882
buy	6689
cart	7323
pv	9706

从 pv 到 buy，转化率为 68.92%,远远高于上一层行为漏斗的 2.27%，因此假设 1 不成立。
#
#假设 2：用户加购/收藏到购买的转化率高
#创建视图，统计不同用户不同商品行为:比较用户从浏览到购买与从收藏/加购到购买的转

化率

```
create view buy_way as
select user_id,item_id,
sum(if(behavior_type='pv',1,0)) as pview,
sum(if(behavior_type='cart',1,0)) as cart,
sum(if(behavior_type='fav',1,0)) as favor,
sum(if(behavior_type='buy',1,0)) as buy
from tb group by user_id,item_id;

#a.浏览总行为数
select count(*) as '总浏览量' from buy_way where pview>0;

#b.浏览到流失
select count(*) as '浏览-流失'
from buy_way
where pview>0 and cart=0 and favor=0 and buy=0;

#浏览-收藏行为计数
select count(*) as '浏览-收藏' from buy_way
where pview>0 and favor>0 and cart=0;

#浏览-加购行为
select count(*) as '浏览-加购' from buy_way
where pview>0 and cart>0 and favor=0;

#浏览-购买行为
select count(*) as '浏览-购买' from buy_way
where pview>0 and buy>0 and favor=0 and cart=0;
#流失
#浏览-收藏-流失
select count(*) as '浏览-收藏-流失'
from buy_way
where pview>0 and favor>0 and cart=0 and buy=0;
#浏览-加购-流失
select count(*) as '浏览-加购-流失'
from buy_way
where pview>0 and favor=0 and cart>0 and buy=0;

#浏览-收藏-购买
select count(*) as '浏览-收藏-购买'
from buy_way
where pview>0 and fav>0 and buy>0 and cart=0;

#浏览-加购-购买
```

```
select count(*) as '浏览-加购-购买'
from buy_way
where pview>0 and fav=0 and buy>0 and cart>0;

#浏览-收藏&加购
select count(*) as '浏览-收藏&加购'
from buy_way
where pview>0 and fav>0    and cart>0;

#浏览-收藏&加购-购买
select count(*) as '浏览-收藏&加购-购买'
from buy_way
where pview>0 and fav>0 and buy>0 and cart>0;
#浏览-收藏&加购-流失
select count(*) as '浏览-收藏&加购-流失'
from buy_way
where pview>0 and fav>0 and buy=0 and cart>0;
```

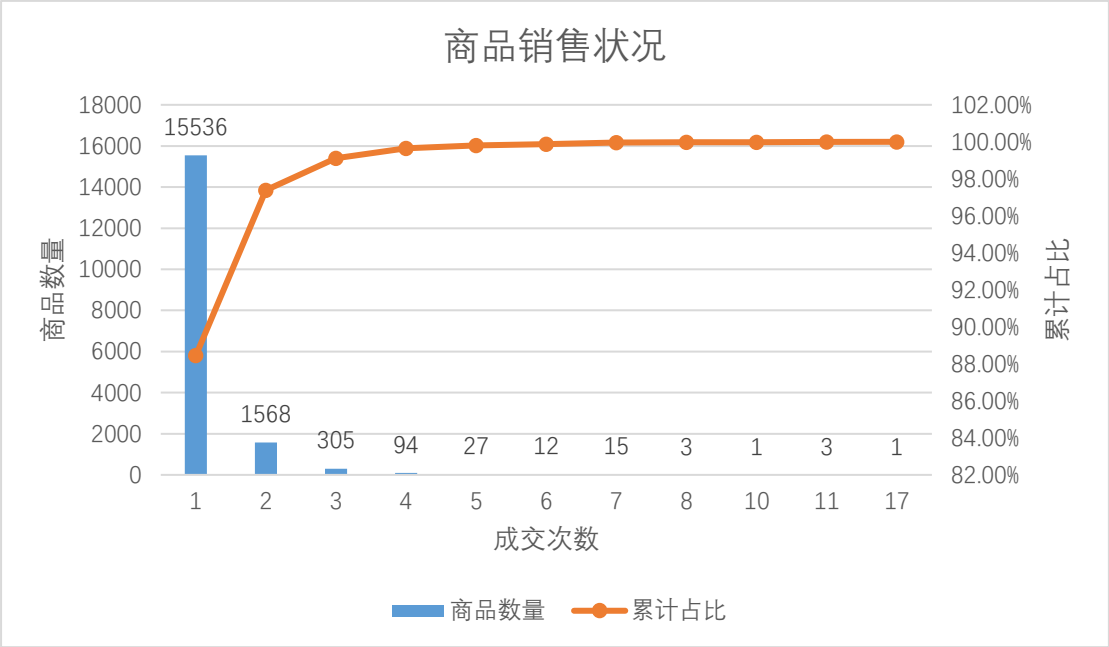


从上图可知，93.4%的用户点击后便流失了，浏览后直接购买的占比仅为 1.3%，浏览后加购的比例则有 3.65%，高于浏览后收藏的比例（1.52%）。从购买转化率来看，收藏的购买转化率为 8.2%，加购的购买转化率更高，为 10.33%，说明用户收藏的商品可能只是喜欢，但还没有很大的购买意向。从这里，可以针对性做一些措施增加用户的购买意向，如发放优惠券或者给予商品一定折扣优惠。收藏并加购的购买转化率比较高，达到 16.92%，这里可以做一些营销手段引导用户收藏并加购。

#5.4 消费者偏好分析

#按商品分组计算其成交次数
select item_id,count(*) as tra_amou
from tb
where behavior_type='buy'
group by item_id order by tra_amou desc;

#按成交次数分组计算商品数量(如成交 x 次的商品有多少个)
select t.tra_amou,count(*)
from (select item_id,count(*) as tra_amou
from tb
where behavior_type='buy'
group by item_id) t
group by t.tra_amou order by t.tra_amou;



从图可知，15536 个商品的成交次数仅为 1 次，占比为 88.45%的，说明了大部分的成交商品为长尾商品，并没有看出特别带动销量的爆款。为了提高销量，商品页可以将畅销品和非畅销品展示在一起，或者捆绑销售，或者将爆款产品集中推送，提高整个平台的销量。

##筛选销量前 10 的商品
select item_id,
sum(if(behavior_type='buy',1,0)) as '商品成交次数',
sum(if(behavior_type='pv',1,0)) as '商品点击量',
sum(if(behavior_type='fav',1,0)) as '商品收藏量',
sum(if(behavior_type='cart',1,0)) as '商品加购量',
concat(round(sum(if(behavior_type='buy',1,0))/sum(if(behavior_type='pv',1,0)),3)*100,'%')
as '商品点击转化率' from tb
group by item_id order by 商品成交次数 desc;

item_id	商品成交次数	商品点击量	商品收藏量	商品加购量	商品点击转化率
3122135	17	16	1	7	106.300%
2964774	11	81	0	7	13.600%
3237415	11	42	1	4	26.200%
2124040	11	4	0	0	275.000%
4401268	10	29	1	0	34.500%
3991727	8	18	1	2	44.400%
1910706	8	12	0	1	66.700%
1004046	8	13	0	4	61.500%
11517	7	32	0	5	21.900%
1595279	7	34	0	8	20.600%

由于数据样本不够大，商品点击率略有不合理，但是不影响后续分析。

#卖的好的商品的原因？

- 假设 1：流量多
- 假设 2：转化率高？收藏量高？加购量高？
- 假设 3：复购率高？

#（1）流量高？
按商品浏览量排序

```
select item_id,
sum(if(behavior_type='buy',1,0)) as '商品成交次数',
sum(if(behavior_type='pv',1,0)) as '商品点击量',
sum(if(behavior_type='fav',1,0)) as '商品收藏量',
sum(if(behavior_type='cart',1,0)) as '商品加购量',
concat(round(sum(if(behavior_type='buy',1,0))/sum(if(behavior_type='pv',1,0)),3)*100,'%') as '商品点击转化率'
from tb
group by item_id order by 商品点击量 desc;
```

item_id	商品成交次数	商品点击量	商品收藏量	商品加购量	商品点击转化率
812879	1	285	5	13	0.400%
3845720	0	222	3	4	0.000%
138964	1	221	4	6	0.500%
3708121	1	193	2	5	0.500%
2032668	2	191	6	7	1.000%
2331370	0	186	7	14	0.000%
2338453	3	180	4	9	1.700%
1535294	7	169	5	20	4.100%
4211339	0	166	3	8	0.000%
3371523	0	161	0	1	0.000%

从商品流量角度来看，浏览量高的商品并未出现在销量最高的商品之中，因此可见商品流量并非商品购买的主要因素，因此假设 1 不成立。

#（2）转化率高？即收藏量 or 加购量高？
#收藏量高？

```
select item_id,
sum(if(behavior_type='buy',1,0)) as '商品成交次数',
sum(if(behavior_type='pv',1,0)) as '商品点击量',
sum(if(behavior_type='fav',1,0)) as '商品收藏量',
sum(if(behavior_type='cart',1,0)) as '商品加购量',
concat(round(sum(if(behavior_type='buy',1,0))/sum(if(behavior_type='pv',1,0)),3)*100,'%') as '商品点击转化率' from tb
group by item_id order by  商品收藏量  desc;
```

item_id	商品成交次数	商品点击量	商品收藏量	商品加购量	商品点击转化率
3330337	1	110	11	7	0.900%
2818406	0	151	11	8	0.000%
2887571	0	72	9	6	0.000%
2783905	1	96	9	4	1.000%
1517532	0	1	9	2	0.000%
2279428	0	125	8	11	0.000%
2364679	0	101	8	10	0.000%
600756	0	52	8	2	0.000%
2778083	0	106	8	4	0.000%
3159978	0	114	8	5	0.000%

显然，收藏量高的商品并未出现在销量最高商品当中，侧面印证了商品收藏后的购买转化率较低；

```
#加购量高？
select item_id,
sum(if(behavior_type='buy',1,0)) as '商品成交次数',
sum(if(behavior_type='pv',1,0)) as '商品点击量',
sum(if(behavior_type='fav',1,0)) as '商品收藏量',
sum(if(behavior_type='cart',1,0)) as '商品加购量',
concat(round(sum(if(behavior_type='buy',1,0))/sum(if(behavior_type='pv',1,0)),3)*100,'%')
as   '商品点击转化率'
from tb
group by item_id order by  商品加购量  desc;
```

item_id	商品成交次数	商品点击量	商品收藏量	商品加购量	商品点击转化率
1535294	7	169	5	20	4.100%
2331370	0	186	7	14	0.000%
3031354	7	159	0	14	4.400%
812879	1	285	5	13	0.400%
1636256	3	25	1	13	12.000%
2279428	0	125	8	11	0.000%
4091349	3	116	6	11	2.600%
2402579	3	37	3	11	8.100%
4649427	0	130	2	10	0.000%
1402604	2	89	3	10	2.200%

商品加购量高的有两个商品成交次数为 7，属于并排销量前 10 的商品，说明用户加购后的

购买转化率高于收藏后的购买转化率。

#（3）复购率高？
#筛选高平均成交量的 top10 商品
```
select item_id,count(distinct user_id) as '成交用户数',
sum(if(behavior_type='buy',1,0)) as '商品成交次数',
sum(if(behavior_type='buy',1,0))/count(distinct user_id) as '平均用户成交量'
from tb where behavior_type='buy'
group by item_id order by 平均用户成交量 desc;
```

item_id	成交用户数	商品成交次数	平均用户成交量
2124040	1	11	11.0000
4296993	1	7	7.0000
1180858	1	7	7.0000
3551756	1	7	7.0000
3953938	1	6	6.0000
4157341	1	6	6.0000
2816569	1	6	6.0000
4792038	1	5	5.0000
1022848	1	5	5.0000
1542908	1	5	5.0000

显然，有几个商品的与销量高商品重合，所以私以为假设 3 成立。由图可知，个别商品成交用户数为 1，但成交次数达到 11 次，可见老客户复购率的提升能有效提高销量和营收。

附：

由于 100 万样本仍有些不足，因此在商品偏好分析这块增加到 200 万样本来做一下对比，得出的结论和 100 万样本相似。

item_id	商品成交次数	商品点击量	商品收藏量	商品加购量	商品点击转化率
3122135	35	39	2	11	89.700%
3237415	17	86	2	4	19.800%
1910706	16	30	1	1	53.300%
2560262	16	234	5	26	6.800%
3031354	15	389	3	39	3.900%
2964774	15	158	0	18	9.500%
705557	14	243	4	23	5.800%
4157341	14	2	0	0	700.000%
1034594	14	7	0	0	200.000%
1595279	13	65	1	12	20.000%

item_id	商品成交次数	商品点击量	商品收藏量	商品加购量	商品点击转化率
2279428	0	273	22	24	0.000%
2818406	3	303	19	18	1.000%
2308741	0	1	17	0	0.000%
2364679	0	255	16	17	0.000%
2331370	3	415	16	25	0.700%
1535294	12	348	16	26	3.400%
1419997	1	130	16	0	0.800%
2453685	4	290	15	17	1.400%
600756	0	102	15	5	0.000%
812879	2	593	14	28	0.300%

item_id	商品成交次数	商品点击量	商品收藏量	商品加购量	商品点击转化率
3031354	15	389	3	39	3.900%
812879	2	593	14	28	0.300%
2560262	16	234	5	26	6.800%
1535294	12	348	16	26	3.400%
2331370	3	415	16	25	0.700%
2279428	0	273	22	24	0.000%
705557	14	243	4	23	5.800%
1684440	8	145	8	22	5.500%
1402604	4	186	5	22	2.200%
4649427	0	271	8	22	0.000%

item_id	成交用户数	商品成交次数	平均用户成交量
4574184	1	13	13.0000
3894457	1	12	12.0000
2124040	1	11	11.0000
3551756	1	7	7.0000
1180858	1	7	7.0000
4157341	2	14	7.0000
3953938	1	6	6.0000
2816569	1	6	6.0000
4462545	1	5	5.0000
1542908	1	5	5.0000