

## 1. 研究背景

本数据集共有 1048575 条数据，其中包含了 2017 年 11 月 25 日至 2017 年 12 月 3 日之间，随机用户的所有行为数据（行为包括点击、购买、加购、喜欢）。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、行为类型和时间戳组成，并以逗号分隔。数据来源：

<https://tianchi.aliyun.com/dataset/dataDetail?dataId=649&userId=1&lang=en-us>

## 2. 分析目的

通过对这些数据进行统计、分析，从中发现用户使用产品的规律，并将这些规律与网站的营销策略、产品功能、运营策略相结合，发现营销、产品和运营中可能存在的问题，解决这些问题就能优化用户体验、实现更精细和精准的运营与营销，让产品获得更好的增长。

## 3. 研究思路

（一）常见用户行为分析模型：

行为事件分析、用户留存分析、漏斗分析、用户行为路径分析、用户分群、点击分析。

（1）电商数据分析基本指标体系

总体运营指标	
流量指标	独立访客数（UV）
	页面访问数（PV）
	人均页面访问数
订单效率指标	总订单数量(buy)
	访问到下单转化率（buy/pv）
总体销售业绩指标	成交额（GMV）
	销售金额
	客单价（成交额/顾客数）
整体指标	销售毛利（销售额-成本）
	毛利率（毛利/销售额）

（2）RFM 模型根据客户活跃程度和交易金额的贡献，进行客户价值细分的一种方法。

R（Recency）——最近一次交易时间间隔。基于最近一次交易日期计算的得分，距离当前日期越近，得分越高。如 5 分制。反映客户交易活跃度。

F（Frequency）——客户在最近一段时间内交易次数。基于交易频率计算的得分，交易频率越高，得分越高。如 5 分制。反映客户交易活跃度。

M（Monetary）——客户最近一段时间内交易金额。基于交易金额计算的得分，交易金额越高，得分越高。如 5 分制。反映客户价值。

RFM 总分值： $RFM=RS \times 100+FS \times 10+MS \times 1$

RFM 分析的主要作用：识别优质客户。可以制定个性化的沟通和营销服务，为更多的营销决策提供有力支持。能够衡量客户价值和客户利润创收能力。

（3）行为事件分析模型：

行为事件分析法具有强大的筛选、分组和聚合能力，逻辑清晰且使用简单，

已被广泛应用。行为事件分析法一般包含事件定义与选择、下钻分析、解释与结论等环节。事件定义与选择。事件描述的是，一个用户在某个时间点、某个地方、以某种方式完成了某个具体的事情。Who、When、Where、What、How 是定义一个事件的关键因素。

其中：Who 是参与事件的主体，对于未登陆用户，可以是 Cookie、设备 ID 等匿名 ID；对于登录用户，可以使用后台配置的实际用户 ID；When 是事件发生的实际时间，应该记录精确到毫秒的事件发生时间；Where 即事件发生的地点，可以通过 IP 来解析用户所在省市；也可以根据 GPS 定位方式获取地理位置信息。How 即用户从事这个事件的方式。用户使用的设备、浏览器、APP 版本、渠道来源等等；What 描述用户所做的这个事件的所有具体内容。比如对于“购买”类型的事件，则可能需要记录的字段有：商品名称、商品类型、购买数量、购买金额、付款方式等。

多维度下钻分析。最为高效的行为事件分析要支持任意下钻分析和精细化条件筛选。当行为事件分析合理配置追踪事件和属性，可以激发出事件分析的强大潜能，为企业回答关于变化趋势、维度对比等等各种细分问题。同时，还可以通过添加筛选条件，可以精细化查看符合某些具体条件的事件数据。

解释与结论。此环节要对分析结果进行合理的理论解释，判断数据分析结果是否与预期相符，如判断产品的细节优化是否提升了触发用户数。如果相悖，则应该针对不足的部分进行再分析与实证。

参考电商数据分析基本指标体系、AARRR 漏斗模型和 RFM 模型，结合适用指标进行分析：

#### （4）用户留存分析：

留存分析是一种用来分析用户参与情况/活跃程度的分析模型，考察进行初始行为的用户中，有多少人会进行后续行为。这是用来衡量产品对用户价值高低的重要方法。

留存分析可以帮助回答以下问题：

一个新客户在未来的一段时间内是否完成了您期许用户完成的行为？如支付订单等；

某个社交产品改进了新注册用户的引导流程，期待改善用户注册后的参与程度，如何验证？

想判断某项产品改动是否奏效，如新增了一个邀请好友的功能，观察是否有人因新增功能而多使用产品几个月？

#### （5）AARRR 漏斗模型：

就是从起点到终点有多个环节，每个环节都会产生用户流失，依次递减，每一步都会有一个转化率。另外衍生出“路径分析方法”，包括：关键路径、扩散路径、收敛路径、端点路径，每一条路径，都是一个漏斗。

#### （6）用户路径分析：

用户在 APP 或网站中的访问行为路径。为了衡量网站优化的效果或营销推广的效果，以及了解用户行为偏好，时常要对访问路径的转换数据进行分析。

以电商为例，买家从登录网站 / APP 到支付成功要经过首页浏览、搜索商品、加入购物车、提交订单、支付订单等过程。而在用户真实的选购过程是一个交缠反复的过程，例如提交订单后，用户可能会返回首页继续搜索商品，也可能去取消订单，每一个路径背后都有不同的动机。与其他分析模型配合进行深入分析后，能为找到快速用户动机，从而引领用户走向最优路径或者期望中的路径。

### （7）用户分群分析：

由于群体特征不同，行为会有很大差别，因此运营人员或者产品人员希望可以根据历史数据将用户进行划分，将具有一定规律特性的用户群体进行归类，进而再次观察该群体的具体行为。这就是用户分群的原理。

### （8）点击分析：

点击图是点击分析方法的效果呈现，在用户行为分析领域，点击分析被应用于显示页面或页面组（结构相同的页面，如商品详情页、官网博客等）区域中不同元素点击密度的图示。包括元素被点击的次数、占比、发生点击的用户列表、按钮的当前与历史内容等因素。

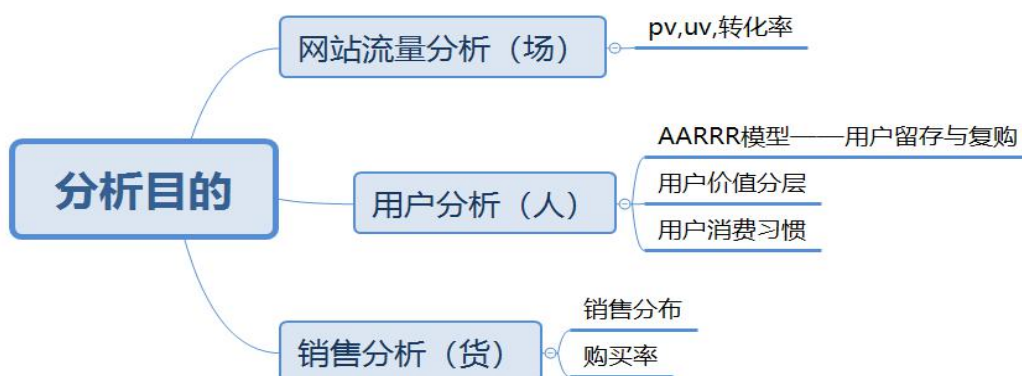
### （二）本文的研究思路

（1）了解网站流量情况；

（2）了解该阶段网站用户粘性以及用户行为习惯；

（3）了解网站商品销售情况；

分析逻辑如下：



[https://blog.csdn.net/light\\_love\\_long](https://blog.csdn.net/light_love_long)

## 4. 分析内容

### 4.1 分析步骤

#### 第一步：理解数据

数据集介绍		
文件名	说明	包含特征
UserBehavior.csv	包含所有的用户行为数据	用户 ID，商品 ID，商品类目 ID，行为类型，时间戳

数据集中每列数据的详细描述	
列名称	说明
User_Id	整数类型，序列化后的用户 ID
Item_Id	整数类型，序列化后的商品 ID
Category_Id	整数类型，序列化后的商品所属类型 ID
Behavior_type	字符串、枚举类型，包括（pv, buy, cart, fav）

Timestamp	行为发生的时间戳
-----------	----------

用户行为类型	说明
pv	商品详情页 pv，等价于点击
buy	商品购买
cart	将商品加入购物车
fav	收藏商品

## 第二步：提出问题

1. 期间用户整体行为情况是怎样？总的点击量，点击人数，购买人数如何？
2. 跳失率如何？用户漏斗分析是怎样的？
3. 每天不同时间段的点击量，点击人数，购买人数情况是怎么样的？
4. 收藏和加购的转化率是怎样的？用户的购买路径分析？
5. 根据 RFM 模型，找出不同价值的用户？

## 4.2 数据清洗

数据清洗（尽量不要再数据库对数据进行增改删）

使用 pandas 进行数据清洗

### 1) 缺失值处理

```
##### 查询缺失值情况#####
data.isnull().sum()
#时间戳列有1个缺失值，查看缺失值列
data[data.iloc[:,4].isnull()]
#3835330行数据缺失
#时间戳缺失值列，用户行为为'P'，这个数据也是异常。查看数据集中用户行为种类
data.iloc[:,3].unique()

#用户行为每个种类有多少数据
data.iloc[:,3].value_counts()

#缺失值只有1列，直接删除。并重置索引
data.dropna(axis=0,inplace=True)
data.reset_index(drop=True,inplace=True)
```

### 2) 重复值处理

由于表为用户行为表，同一个用户可以有多个行为动作，且可以相同，所以数据可以有重复。

```
##### 重复数据处理#####
#查看重复数据
data[data.duplicated()]
#删除重复数据
data.drop_duplicates(inplace=True)
#充值索引
data.reset_index(drop=True,inplace=True)
#数据清洗完成，导出到本地
data.to_excel(r'E:\实习\userbehavior\UserBehavio_Done.xlsx',sheet_name='data1',startcol=0,index=False)
data.to_csv(r'E:\实习\userbehavior\UserBehavio_Done.csv',index=False)
```

### 3) 数据集选择

```
##时间戳转换为日期、时间数据。并把日期和时间分为两列
##直接转时间格式失败，需要借助Lambda
data.loc[:, 'Timestamp'] = data['Timestamp'].apply(lambda x: time.strftime('%Y-%m-%d %H:%M:%S', time.localtime(x)))
data.loc[:, 'Date'] = data['Timestamp'].apply(lambda x: x.split(' ')[0]) #日期列
data.loc[:, 'Time'] = data['Timestamp'].apply(lambda x: x.split(' ')[1]) #时间列
##将数据转换成时间格式
data['Date'] = pd.to_datetime(data['Date']) #将Date列，由str转换为时间格式
data['Time'] = pd.to_datetime(data['Time']) #将Date列，由str转换为时间格式

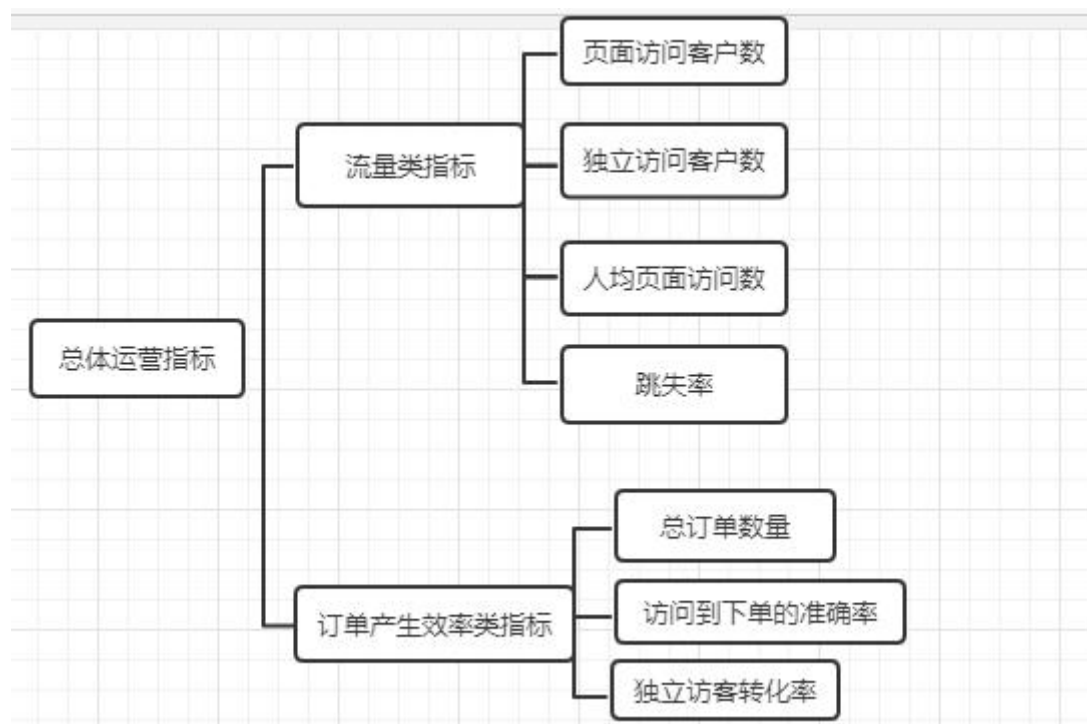
##提取周，小时
data['hour'] = data['Time'].dt.hour #小时列
##把时间戳列删除
data = data.drop(columns='Timestamp', axis=1)

##### 异常值处理#####
##原始数据日期区间为2017-11-25到2017-12-03，这个时间区间外认为是异常数据。查看一下具体情况
data[data['Date'] < '2017-11-25'].shape
data[data['Date'] < '2017-11-25']['Date'].value_counts()
data[data['Date'] > '2017-12-03'].shape
data[data['Date'] > '2017-12-03']['Date'].value_counts()
data['Date'].head()
##时间区间外的日期数据删除
data = data[(data['Date'] >='2017-11-25') & (data['Date'] <='2017-12-03')]
```

## 4.3 构建模型

### (一) 总体运营分析

用户整体行为情况是怎样？总的点击量，点击人数，购买人数如何？



(1) pv（总访课数）：940007

```
-- 1.计算浏览量: pv
SELECT behavetype, COUNT(*)
FROM USER
WHERE behavetype='pv';
```

(2) uv（用户总数）：10207

```
-- 4.计算独立访客数uv
SELECT COUNT(DISTINCT user_id)
FROM USER;
-- 5.计算有购买行为的访客数
```

(3) 购买人数：7004

```
-- 5.计算有购买行为的访客数
SELECT COUNT(DISTINCT user_id)AS 购买用户量
FROM USER
WHERE behavetype='buy';
```

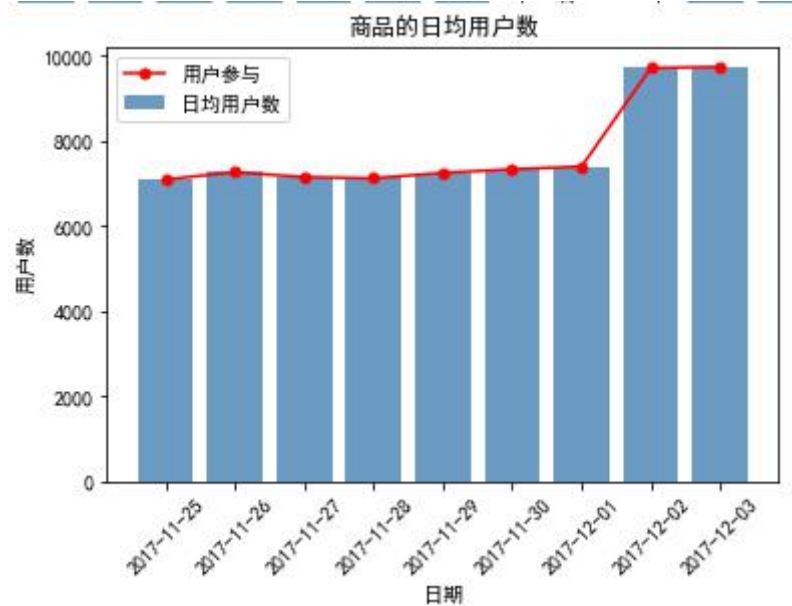
(4) 复购率：66.28%

```
-- 4.复购率：产生两次或两次以上购买的用户占购买用户的比例:66.28%
SELECT ((SELECT COUNT(DISTINCT user_id) FROM USER
WHERE user_id IN (SELECT user_id FROM USER WHERE behavetype='buy' GROUP BY user_id HAVING COUNT(user_id)>=2))
/(SELECT COUNT(DISTINCT user_id)AS 用户数 FROM USER WHERE behavetype='buy'))AS 复购率;
```

(5) 日均访问量

```
-- 5.计算日均访客数
SELECT dates,COUNT(DISTINCT user_id)
FROM USER
WHERE behavetype='pv'
GROUP BY dates
ORDER BY dates;
```

dates	日均访客数
2017-11-25	7099
2017-11-26	7270
2017-11-27	7155
2017-11-28	7127
2017-11-29	7255
2017-11-30	7339
2017-12-01	7403
2017-12-02	9717
2017-12-03	9744





分析：11月25日至12月1日使用用户数变化幅度小，范围在7000到8000之间，从12月2日有较大幅度增加，对比11月25，26（周末）有较大的提升，可能是周末的双12预热营销活动效果。

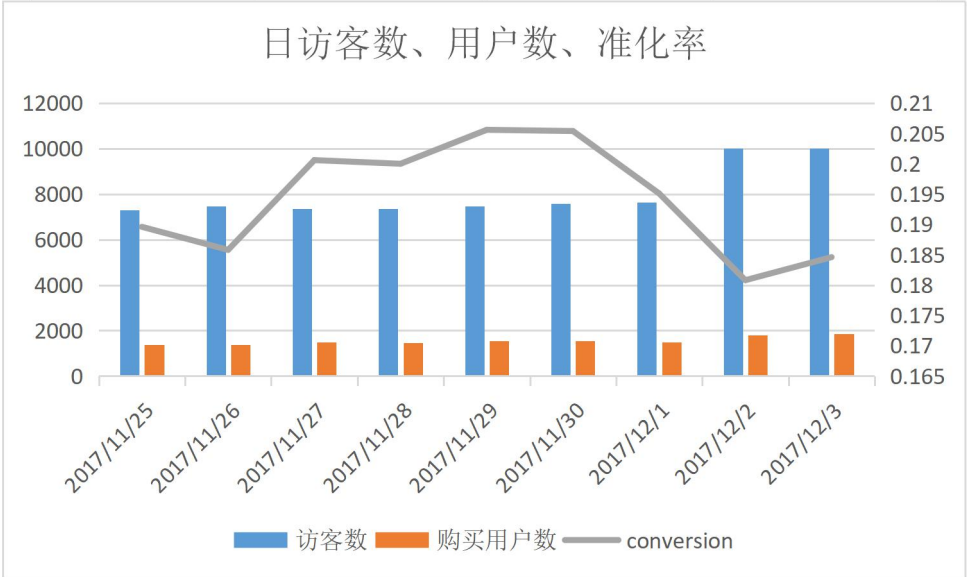
(6)人均页面访客数：92.09

```
-- 6.人均页面访客数
SELECT CAST((SELECT COUNT(*) FROM USER WHERE behavetype='pv')/COUNT(DISTINCT user_id)AS DECIMAL(10,2))AS 'pv/uv'
FROM USER;
```

(6)网站购买转化率（bv/uv）（支付转化率=支付访客数/总访客数）

```
-- 7.转化率=（产生购买行为的客户人数 / 所有到达店铺的访客人数）* 100%。
SELECT t1.dates,t1.bv AS 购买用户数,t2.uv AS 访客数 ,CONCAT(ROUND(t1.bv/t2.uv*100,2),'%') conversion
FROM(SELECT dates,COUNT(DISTINCT user_id) bv FROM USER WHERE behavetype='buy' GROUP BY dates)AS t1
LEFT JOIN
(SELECT dates,COUNT(DISTINCT user_id) uv FROM USER GROUP BY dates)AS t2
ON t1.dates=t2.dates
ORDER BY dates;
```

dates	访客数	购买用户数	conversion
2017-11-25	7304	1385	18.96%
2017-11-26	7477	1389	18.58%
2017-11-27	7363	1477	20.06%
2017-11-28	7359	1472	20.00%
2017-11-29	7475	1537	20.56%
2017-11-30	7580	1557	20.54%
2017-12-01	7627	1488	19.51%
2017-12-02	10024	1812	18.08%
2017-12-03	10021	1850	18.46%



访客数在11月25-12月1日基本稳定，保持在7000-8000之间，可能由于双12活动的原因，12月2日开始访客数增加到10000左右。

11月25-12月1日期间访客数基本稳定，保持在1500左右，12月2日开始，购买用户数开始增加到了1800左右。

转化率在11月30日达到峰值，之后呈下降趋势，在12月2日的转化率最低，如何开始又有上升趋势。

(二) 用户漏斗分析

(1) 跳失率:5.76%

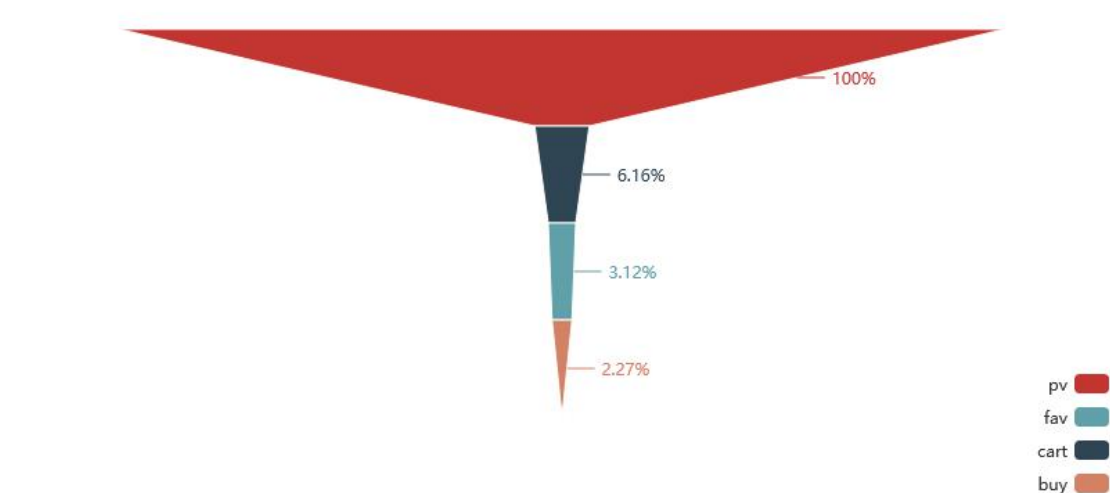
```
-- 1.计算跳失率
]SELECT CONCAT(ROUND((SELECT COUNT(DISTINCT user_id)
FROM USER
WHERE user_id NOT IN(SELECT user_id FROM USER WHERE behavetype='buy')
AND user_id NOT IN(SELECT user_id FROM USER WHERE behavetype='cart')
AND user_id NOT IN(SELECT user_id FROM USER WHERE behavetype='fav'))/(SELECT COUNT(DISTINCT user_id) FROM USER )*100,2),'%')A
```

## (2) 用户漏斗分析

```
-- 转化漏斗
SELECT behavetype,COUNT(user_id)
FROM USER
GROUP BY behavetype
ORDER BY COUNT(user_id) DESC;
```

行为类型	点击量	转化率
pv	940007	100%
cart	57889	6.16%
buy	29353	3.12%
fav	21326	2.27%

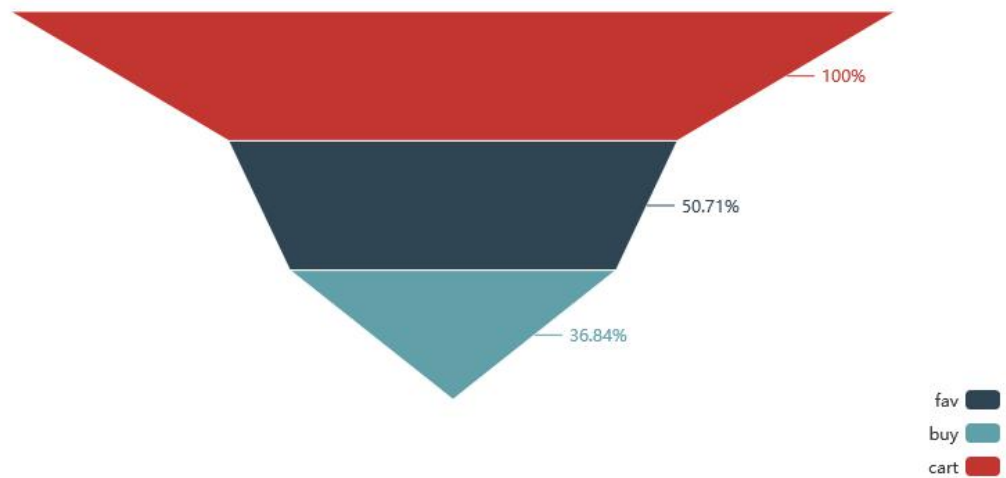
漏斗图



一共有 940007 的浏览量,其中转化为加购的有 57889, 转化为购买的有 29353 条, 转化为收藏的有 21326 条记录。以浏览量为基数, 从浏览到加购的转化率为 6.16%, 从浏览到收藏的转化率为 3.12%, 从浏览到购买的转化率仅有 2.27%。



漏斗图

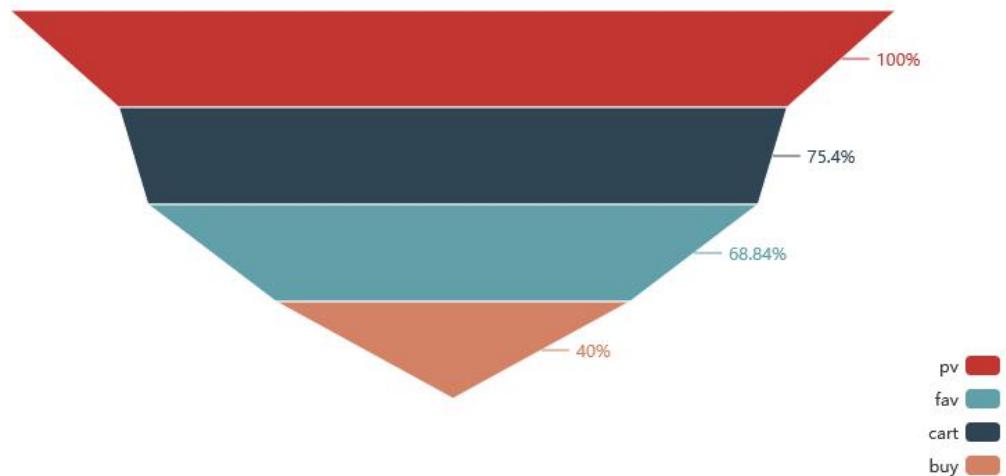


以加购量为基数，从加购到收藏的转化率为 50.71%，从浏览到收藏的转化率为 36.84%。

### (3) 独立访客转化漏斗

```
-- 独立访客转化
SELECT behavetype, COUNT(DISTINCT user_id)
FROM USER
GROUP BY behavetype
ORDER BY COUNT(DISTINCT user_id) DESC;
```

独立访客转化漏斗图



独立访客从浏览到购买的转化率能达到 40%，独立访客从浏览到加购的转化率为 75.4%，独立访客从浏览到收藏的转化率为 68.84%，若要了解具体环节的转化率，需要对整个流程接触到的版面进行细化的数据分析。

### (三) 时间序列分析

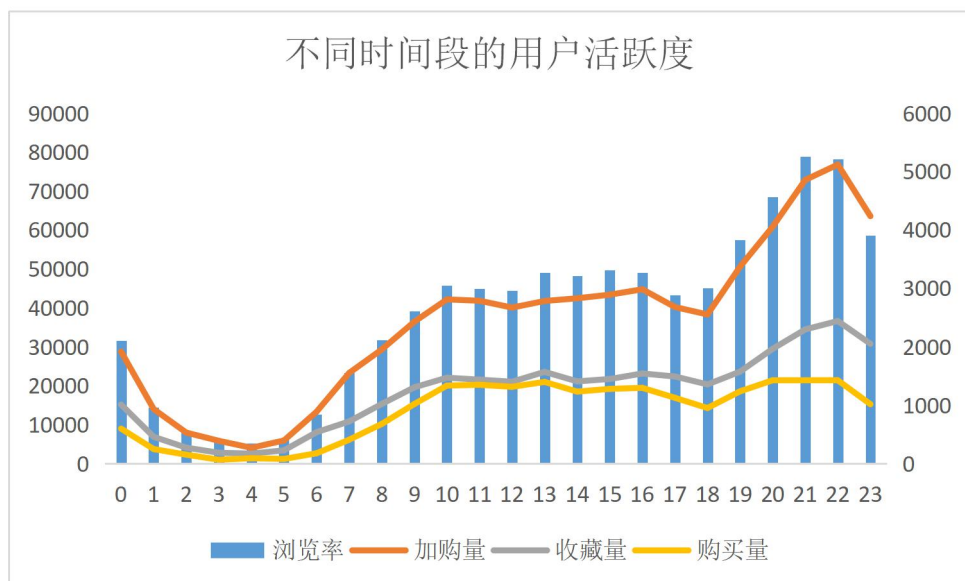
每天不同时间段的点击量，点击人数，购买人数情况

```

-- 1.不同时间段的点击量
SELECT t1.hours,t1.浏览量,t1.加购量,t1.收藏量,t1.购买量,t2.用户数,t3.购买用户数
FROM(SELECT hours,SUM(CASE WHEN behavetype='pv' THEN 1 ELSE 0 END) AS 浏览量,
SUM(CASE WHEN behavetype='cart' THEN 1 ELSE 0 END)AS 加购量,
SUM(CASE WHEN behavetype='fav' THEN 1 ELSE 0 END)AS 收藏量,
SUM(CASE WHEN behavetype='buy' THEN 1 ELSE 0 END)AS 购买量
FROM USER
GROUP BY hours)AS t1
LEFT JOIN
(SELECT hours,COUNT(DISTINCT user_id)AS 用户数
FROM USER
GROUP BY hours)AS t2
ON t1.hours=t2.hours
LEFT JOIN(SELECT hours,COUNT(DISTINCT user_id)AS 购买用户数
FROM USER
WHERE behavetype='buy'
GROUP BY hours)AS t3
ON t1.hours=t3.hours
ORDER BY t1.hours;

```

hours	浏览量	加购量	收藏量	购买量
00	31671	1918	1010	598
01	14511	933	458	248
02	8306	530	272	152
03	5826	389	185	65
04	5146	270	170	91
05	6316	395	226	79
06	12709	887	537	176
07	23404	1551	721	408
08	31758	1953	1020	678
09	39107	2428	1305	1019
10	45716	2811	1470	1335
11	44877	2788	1437	1349
12	44406	2671	1401	1314
13	49040	2784	1570	1398
14	48280	2830	1407	1232
15	49740	2892	1448	1279
16	49117	2984	1542	1296
17	43275	2680	1491	1126
18	45048	2553	1355	951
19	57536	3372	1574	1237
20	68456	4056	1965	1426
21	78955	4857	2297	1427
22	78246	5120	2441	1426
23	58561	4237	2051	1016



从点击量，加购量，收藏量，购买量的时序图来看，四个特质指标的变化趋势相同。从早晨 5 点-10 点用户点击量呈较快增长阶段。到 10 点-16 点，用户点击量在 50000 左右波动。17-18 点，点击量略有下滑。19 点以后点击量进入较快增长阶段，且在 20-21 点达到顶峰，之后点击量开始下降，到凌晨 4 点下降到最低点。

从纵向水平来看，点击量>加购量>收藏量>购买量

网站可以利用这些时间点进行营销活动安排。另外总体来看购买转化率更贴近于收藏率，可以适当利用来进行销售量预测。

（四）用户分析（数据有限，用户新增及激活不做分析）

（1）用户的行为分析

-- 1.各用户的行为

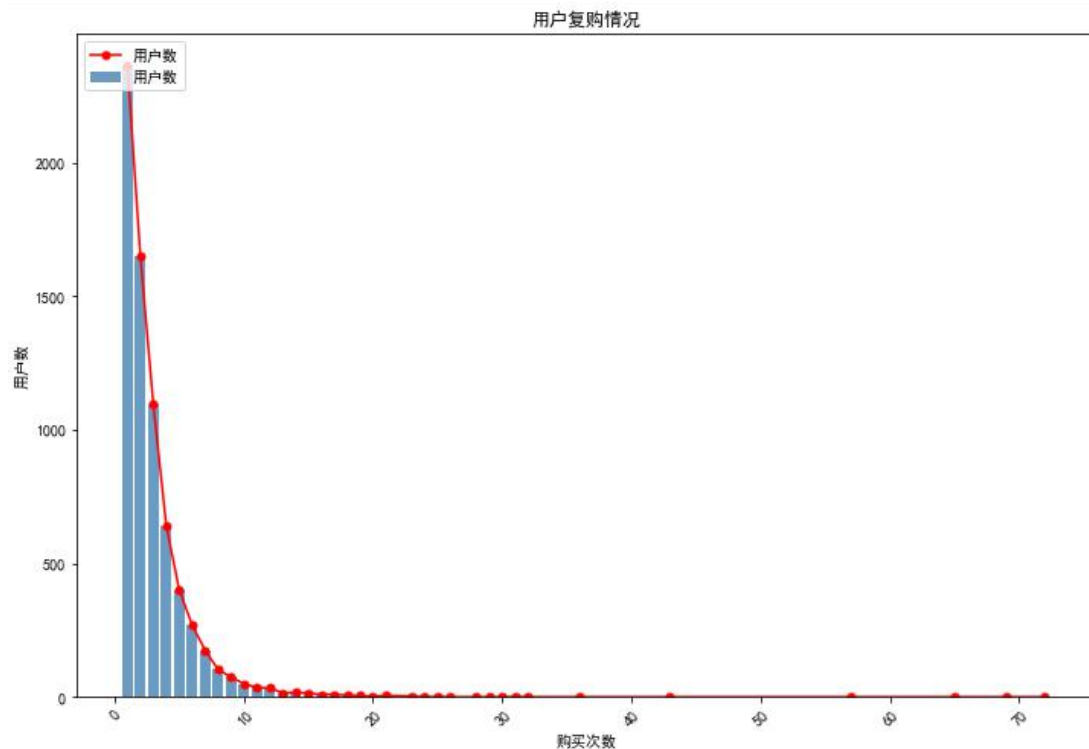
```
SELECT user_id,COUNT(behavetype),
SUM(CASE WHEN behavetype='pv' THEN 1 ELSE 0 END) AS 浏览率,
SUM(CASE WHEN behavetype='cart' THEN 1 ELSE 0 END)AS 加购量,
SUM(CASE WHEN behavetype='fav' THEN 1 ELSE 0 END)AS 收藏量,
SUM(CASE WHEN behavetype='buy' THEN 1 ELSE 0 END)AS 购买量
FROM USER
GROUP BY user_id
ORDER BY COUNT(behavetype) DESC
LIMIT 0,11000;
```

<input checked="" type="checkbox"/>	user_id	count(behavetype)	浏览率	加购量	收藏量	购买量
<input checked="" type="checkbox"/>	115477	781	755	26	0	0
<input checked="" type="checkbox"/>	116139	724	724	0	0	0
<input checked="" type="checkbox"/>	114912	689	616	64	6	3
<input checked="" type="checkbox"/>	115906	668	649	18	0	1
<input checked="" type="checkbox"/>	1010419	665	625	39	1	0
<input checked="" type="checkbox"/>	1014799	607	583	7	17	0
<input checked="" type="checkbox"/>	101249	585	566	10	1	8
<input checked="" type="checkbox"/>	121233	582	564	10	8	0
<input checked="" type="checkbox"/>	119229	582	552	5	24	1
<input checked="" type="checkbox"/>	1006307	576	547	13	4	12
<input checked="" type="checkbox"/>	1015600	560	534	6	18	2

可以挖掘重点活跃对象进行分析，或者分析活跃度低的用户产生的原因，从而实现精准营销。

## (2) 复购分析

```
-- 2.变现—复购分析
SELECT t1.购买次数,COUNT(t1.user_id) 用户数
FROM(SELECT user_id,COUNT(*)AS 购买次数 FROM USER WHERE behavetype='buy'
GROUP BY user_id
ORDER BY COUNT(*) DESC)AS t1
GROUP BY t1.购买次数
ORDER BY t1.购买次数;
```



复购次数集中在 1-5 次，呈长尾分布，该阶段用户消费欲望不大。

(3) 复购率：产生两次或两次以上购买的用户占购买用户的比例:66.28%

```
-- 3.复购率：产生两次或两次以上购买的用户占购买用户的比例:66.28%
]SELECT ((SELECT COUNT(DISTINCT user_id) FROM USER
-WHERE user_id IN (SELECT user_id FROM USER WHERE behavetype='buy' GROUP BY user_id HAVING COUNT(user_id)>=2))
/(SELECT COUNT(DISTINCT user_id)AS 用户数 FROM USER WHERE behavetype='buy'))AS 复购率;
```

## (五) 用户价值分析

借鉴 RFM 模型，暂时不考虑 M（客户最近一段时间内交易金额）维度，对 R（最近一次购买时间），F（购买频率）进行打分，完成用户分层。

```

-- (五)RMF分析
SELECT allrank.user_id,allrank.recent,allrank.buynum,
CONCAT(CASE WHEN allrank.r_r <= (6382)/4 THEN '4'
        WHEN allrank.r_r >(6382)/4 AND allrank.r_r <=(6382)/2 THEN '3'
        WHEN allrank.r_r > (6382)/2 AND allrank.r_r <=(6382)/4*3 THEN '2'
        ELSE '1' END,
CASE WHEN allrank.f_r<=(856)/4 THEN '4'
        WHEN allrank.f_r > (856)/4 AND allrank.f_r <=(856)/2 THEN '3'
        WHEN allrank.f_r > (856)/2 AND allrank.f_r <=(856)/4*3 THEN '2'
        ELSE '1' END)AS user_value
FROM(SELECT recenter_value.user_id,recenter_value.recent,rank()over(ORDER BY recenter_value.recent) r_r,
frequency_value.buynum,rank()over(ORDER BY frequency_value.buynum DESC) f_r
FROM
(SELECT user_id,DATEDIFF('2017-12-04',MAX(dates))AS recent
FROM USER
WHERE behavetype='buy'
GROUP BY user_id)AS recenter_value
LEFT JOIN(SELECT user_id,COUNT(*)AS buynum
FROM USER
WHERE behavetype='buy'
GROUP BY user_id)AS frequency_value
ON recenter_value.user_id=frequency_value.user_id) allrank
ORDER BY user_value DESC;

```

<input checked="" type="checkbox"/>	user_id	recent	buynum	user_value
<input checked="" type="checkbox"/>	107932	1	72	44
<input checked="" type="checkbox"/>	122504	1	69	44
<input checked="" type="checkbox"/>	128379	1	65	44
<input checked="" type="checkbox"/>	1003983	1	43	44
<input checked="" type="checkbox"/>	128113	1	32	44
<input checked="" type="checkbox"/>	114948	1	31	44
<input checked="" type="checkbox"/>	102953	1	29	44
<input checked="" type="checkbox"/>	100101	1	28	44
<input checked="" type="checkbox"/>	1014610	1	25	44
<input checked="" type="checkbox"/>	1007071	1	24	44
<input checked="" type="checkbox"/>	1006872	1	23	44

- 1) 哪一类客户是最优质的客户？
- 2) 哪一类客户是具有高流失风险的优质客户？
- 3) 哪一类客户是需要进行 Upsell 的客户？
- 4) 哪一类客户是公司的高成本客户（不赚钱客户）？

消费进度、消费频度、消费额度是测算消费者价值最重要也是最容易的方法，这充分的表现了这三个指标对营销活动的指导意义。但不代表这三个指标牢不可破，业务方除了消费额度、消费频度以外，在制定补贴策略的时候，还会考虑用户的跨场景使用，越多的功能业务场景被使用，意味着用户忠诚度越高。

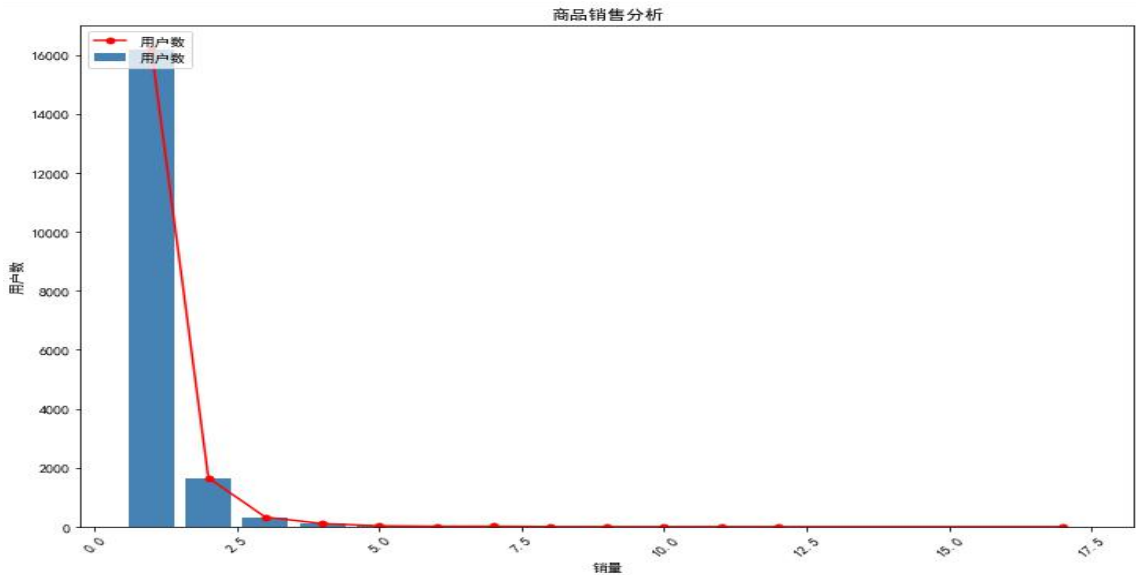
## (六) 商品销售分析

### (1) 商品销售分布

销量	用户数
1	16195
2	1656
3	324
4	106
5	28
6	14
7	15
8	3



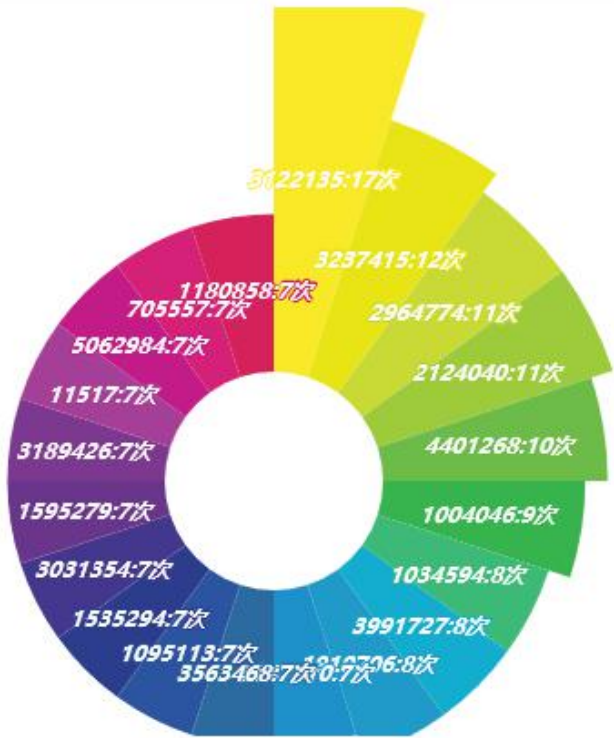
9	1
10	1
11	2
12	1
17	1



大部分销售来源于销售次数为 1 的 16195 个商品，说明该时段销售主要源于长尾部分，而非销售数量较高的畅销品。

(2) 畅销品及其品类分析

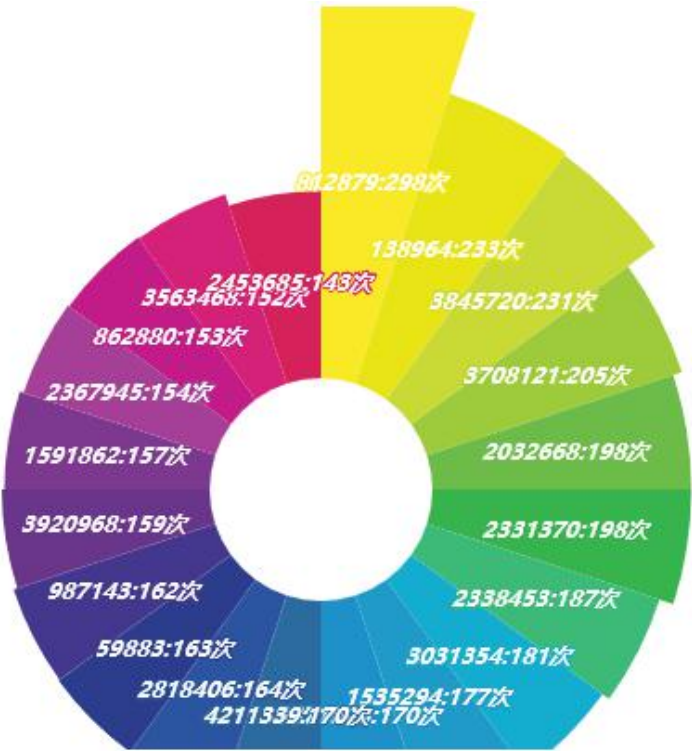
商品销售top20图





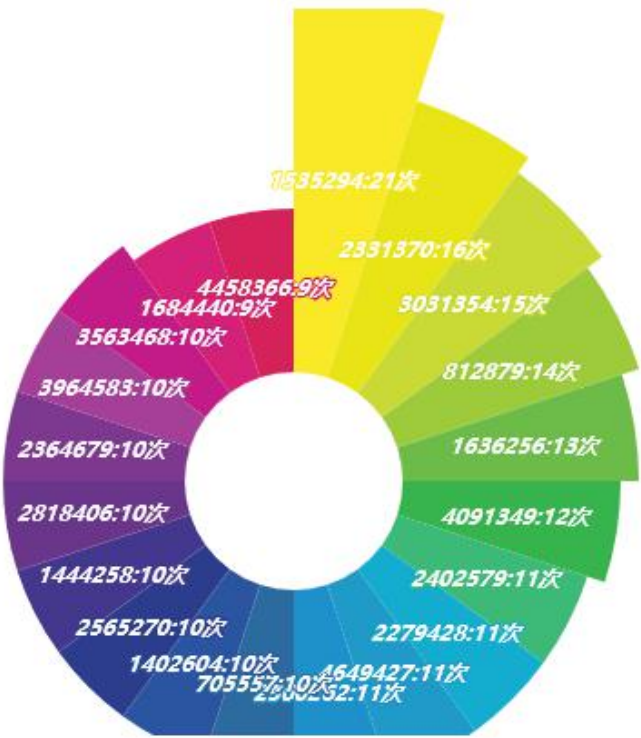
查看销量 top20 的商品及品类分布：销量为前 3 的商品为 3122135 共 17 次，3237415，销量为 12，2964774，销量为 11，分属 1516409，3231625，1265358 类。

商品浏览量top20图



查看浏览 top20 的商品及品类分布：浏览为前 3 的商品为 812879，138964，3845720，分属 4756105，3607361，3607361 类。各品类商品浏览量最高的为 3607361 共浏览次数 849，其次为 4756105，浏览量为 832。1859277，浏览量为 324。

商品加购量top20图



查看加购 top20 的商品及品类分布:加购为前 3 的商品为 1535294,2331370,3031354, 分属 3848953, 3607361, 2735466 类。

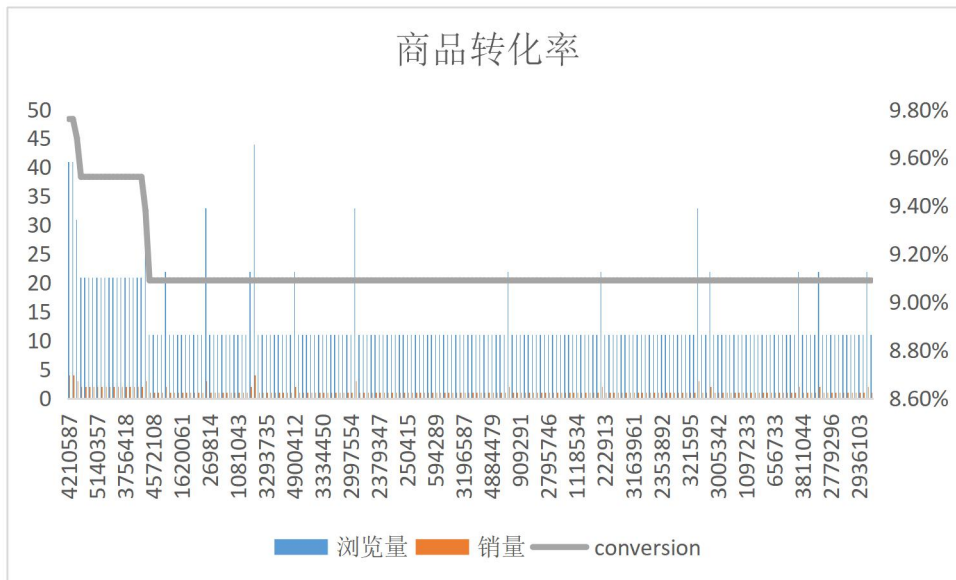
商品收藏量top20图



查看收藏 top20 的商品及品类分布:收藏为前 3 的商品为 2818406,3330337,2308741, 分属 982926, 4756105, 4827153 类

(1) 商品购买转化率

```
-- (七) (3) 商品购买转化率
SELECT t1.category_id,t1.item_id,t1.销量,t2.浏览量,CONCAT(ROUND(t1.销量/t2.浏览量*100,2),'%') conversion
FROM
(SELECT category_id,item_id,COUNT(*)AS 销量
FROM USER
WHERE behavetype='buy'
GROUP BY item_id)AS t1
LEFT JOIN
(SELECT category_id,item_id,COUNT(*)AS 浏览量
FROM USER
WHERE behavetype='pv'
GROUP BY item_id)AS t2
ON t1.item_id=t2.item_id
ORDER BY conversion DESC
LIMIT 200;
```



## 5. 总结

### (1) 关于平台流量与转化率:

整体来看从浏览到有购买意愿有 2.27%的转化率，从有购买意愿到有购买行为有 36.84%的转化率，说明大多数用户以浏览页面为主而购买转化较少，低购买意愿转化率可能与刚完成的双 11 大促有关。从 9 天的趋势来看，前 7 天转化率处于稳步上升阶段，说明双 11 之后的消费情况在回升，进入 12 月之后，浏览量有稍大幅度的提升，加购转化率有小幅提升，购买转化率大幅下降，可能双 12 活动有关。是否需要提高转化率，可以对比去年同时段转化率情况，使用营收目标判断转化率是否达到目标。

提高加购转化率的建议:

- 优化搜索引擎，利用用户画像优化商品匹配;
- 优化商品界面加购与收藏按键布局，以使用户触达。

### 2、关于网站用户粘性以及用户消费习惯

提高用户留存建议:

——分析流失用户特征，进行用户流失预测，结合用户画像，通过各种活动进行用户唤醒。

对于用户复购情况，9 天内网站有复购现象的用户数接近 60%，但是总体上约 30%的用户产生了 80%的消费次数，复购次数多的用户偏少，可能与双 11 刚结束，双 12 未开始的特殊时段有关，建议拉长分析区间分析复购情况。

——提高复购率的建议：开展营销活动，比如淘宝的达成金主的条件限制，鼓励用户复购。

对于用户消费习惯的分析，通过了解用户消费的集中时段集中实施各种营销活动，提高投入产出比。

### 3、关于商品销售情况

分析发现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，这也是双 11 之后用户的补充采买的特征，同时发现此时用户购买的品类以及商品的浏览量很低，用户的个人喜好特征表现明显，同时浏览量高的商品购买转化率低。

——针对浏览量高而购买转化率低的商品，建议商家改善商品页面，详情页以及评论区管理。