# 淘宝用户购物行为数据可视化分析

## 一、项目背景与目的

### 1.1项目背景

UserBehavior是阿里巴巴提供的一个淘宝用户行为数据集，用于隐式反馈推荐问题的研究。数据集包含了2017年11月25日至2017年12月3日之间，有行为的约一百万随机用户的所有行为（行为包括点击、购买、加购、喜欢）。

### 1.2项目目标

流量分析：PV/UV是多少，通过分析PV/UV能发现什么规律？

漏斗分析：用户“浏览-收藏-加购-购买”的转化率是怎样的？哪一步的折损比例最大？

用户价值分析：对电商平台什么样的用户是有价值的？如果你作为商家，要重点关注哪部分用户？

### 1.3数据来源

数据集: 淘宝用户购物行为数据集

<https://tianchi.aliyun.com/dataset/649?spm=a2c22.27080692.J_3941670930.10.7668569939Ijhm&from=search>

UserBehavior是阿里巴巴提供的一个淘宝用户行为数据集，用于隐式反馈推荐问题的研究。

|  |  |  |
| --- | --- | --- |
| 文件名称 | 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 |

|  |
| --- |
| UserBehavior.csv |
| 本数据集包含了2017年11月25日至2017年12月3日之间，有行为的约一百万随机用户的所有行为（行为包括点击、购买、加购、喜欢）。数据集的组织形式和MovieLens-20M类似，即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。 |

|  |  |
| --- | --- |
| 列名称 | 说明 |
| 用户ID | 整数类型，序列化后的用户ID |
| 商品ID | 整数类型，序列化后的商品ID |
| 商品类目ID | 整数类型，序列化后的商品所属类目ID |
| 行为类型 | 字符串，枚举类型，包括('pv', 'buy', 'cart', 'fav') |
| 时间戳 | 行为发生的时间戳 |

注意到，用户行为类型共有四种，它们分别是

|  |  |
| --- | --- |
| 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 |

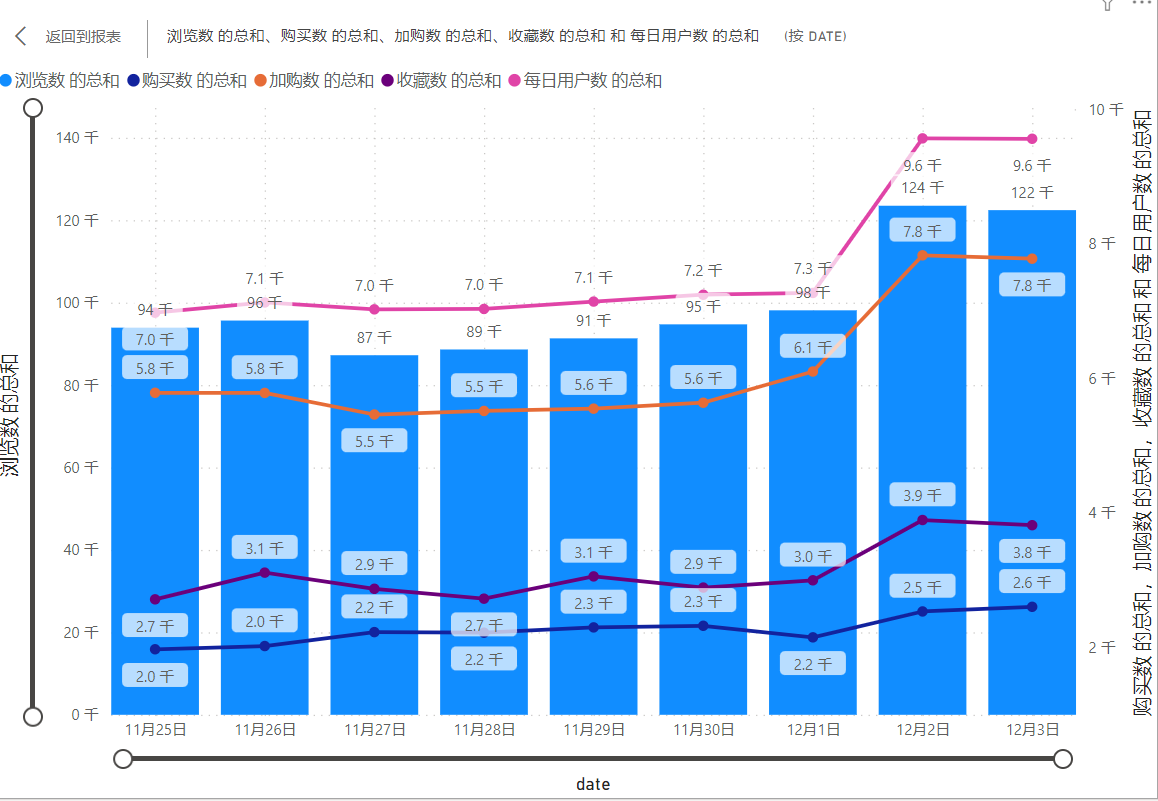
关于数据集大小的一些说明如下

|  |  |
| --- | --- |
| 维度 | 数量 |
| 用户数量 | 987,994 |
| 商品数量 | 4,162,024 |
| 用户数量 | 987,994 |
| 商品类目数量 | 9,439 |
| 所有行为数量 | 100,150,807 |

## 二、流量分析

### 2.1基于天级别访问流量分析





11月25日至12月1日数据平稳，12月2日至3日数据显著增长，且与前一周末无明显关联。日活跃用户、浏览、收藏、加购涨幅超过购买数，推测与淘宝双十二预热活动相关，预热提升了前置购买行为。

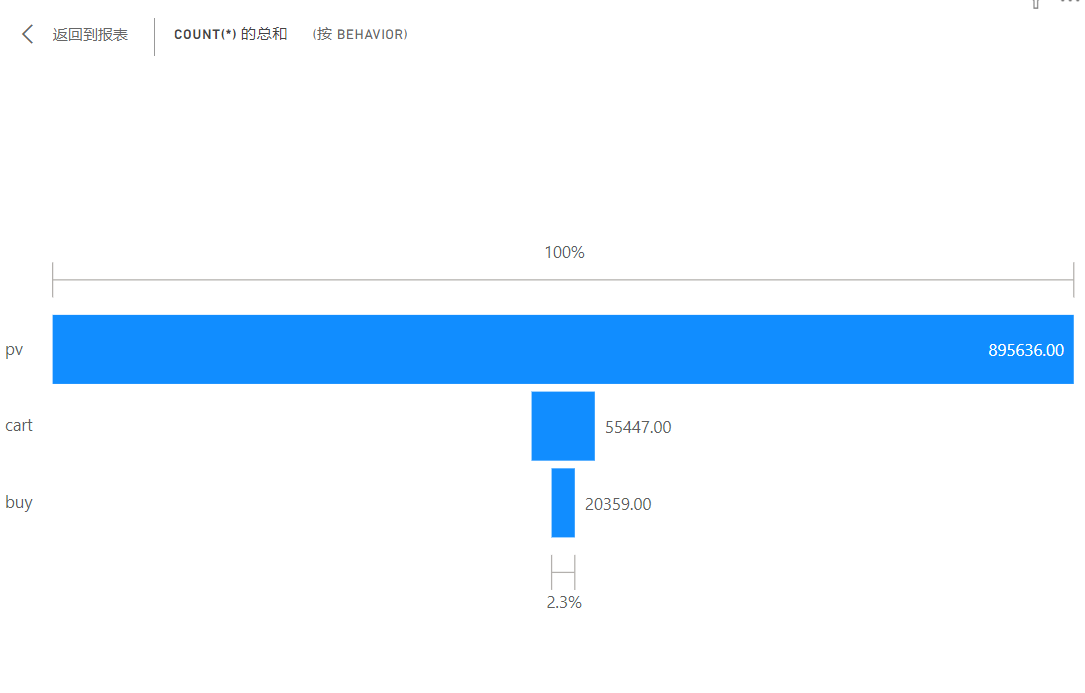
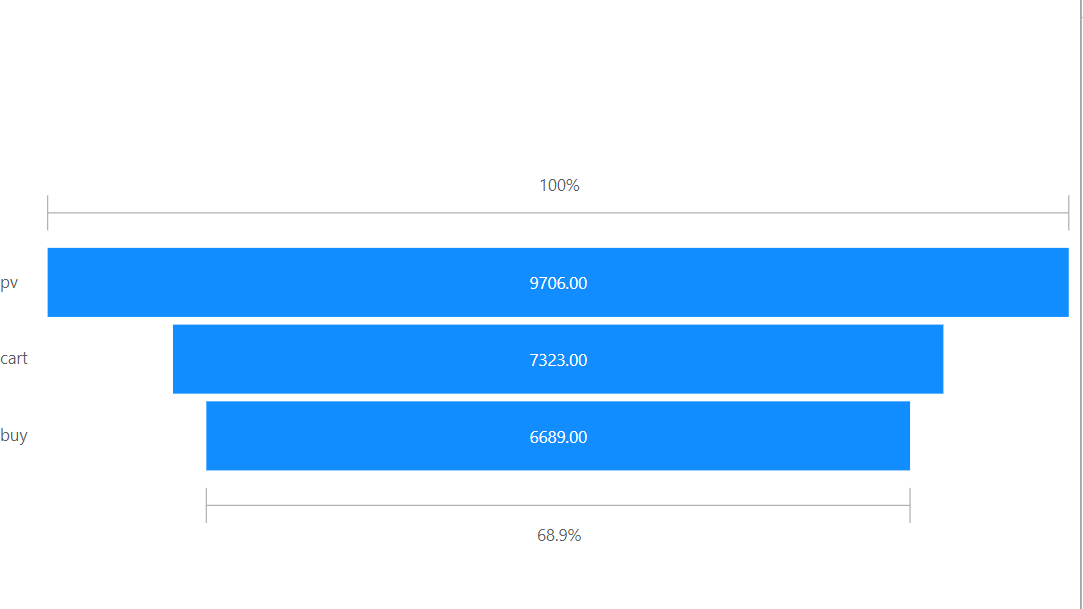
### 2.2基于小时级别访问流量分析

用户行为数据在凌晨2点至5点降至低谷，上午9点至下午6点间数据呈现小高峰，但中午12点和下午4点至5点略有下降。傍晚6点至晚上11点数据大幅上升，可能原因是，淘宝用户群体主要是工薪阶层，这部分群体在下班后开始使用淘宝浏览购物，晚上9点左右达到日峰值。这一趋势与用户日常作息习惯相吻合。

## 三、漏斗分析

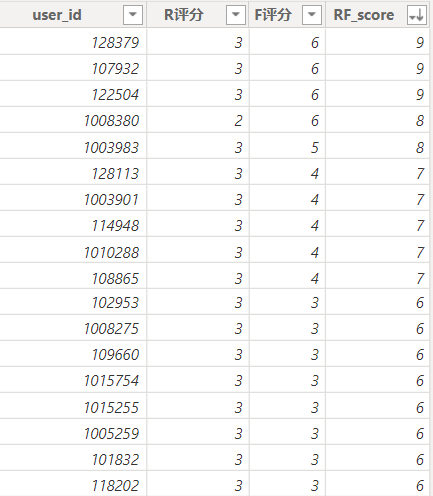
PV/UV约等于92。PV/UV比值较高反映了用户参与度较高，这通常意味着网站内容吸引人且易于导航。清晰的网站结构和高质量的内容也会促使用户浏览更多页面，从而提高这个比值。

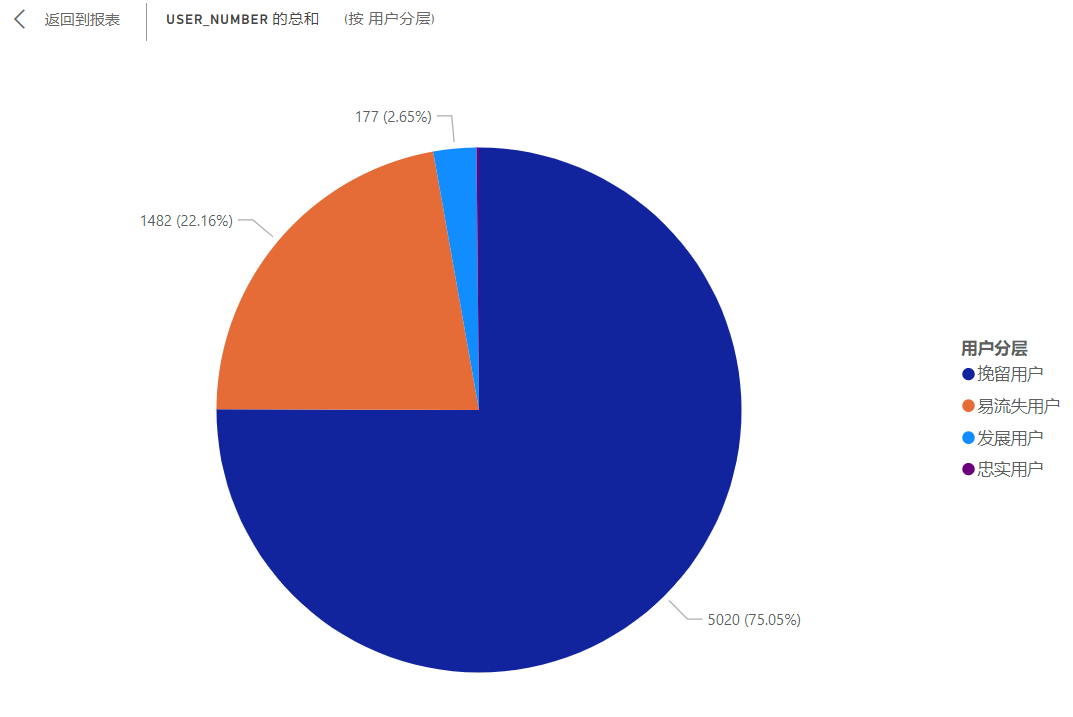




综合用户总行为传化漏斗图和独立访客转化漏斗图可以得出:用户总浏览PV-加购转化率是6.19%，用户加购-购买转化率是36.72%。独立访客UV-加购转化率是75.45%，用户加购-购买转化率是91.34%。从浏览到加购的折损最大。

## 四、用户价值分析



挽留用户占比高，潜力大，应定期促活，如上新提醒、优惠发放和精准推荐，以提高消费频率。

易流失用户占比高，需调研原因并采取价格激励、优惠发放等召回措施。

发展用户占比较低，可推送新品和促销信息来提升消费。

忠诚用户占比低，是高价值用户，需制定专属策略如优惠和客服来保持粘性。