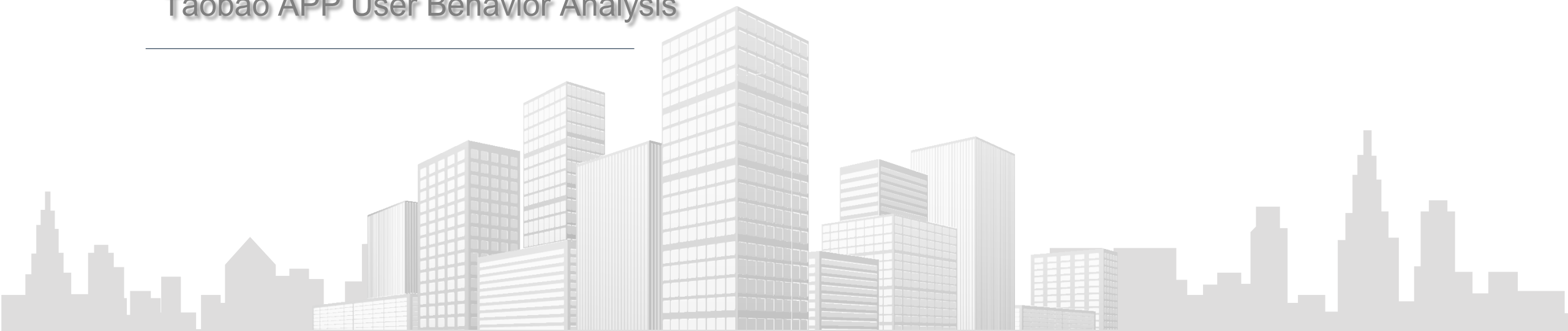




淘宝APP用户行为分析

Taobao APP User Behavior Analysis



CONTENTS



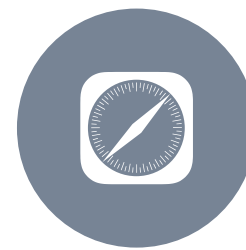
前言



数据来源



用户行为分析



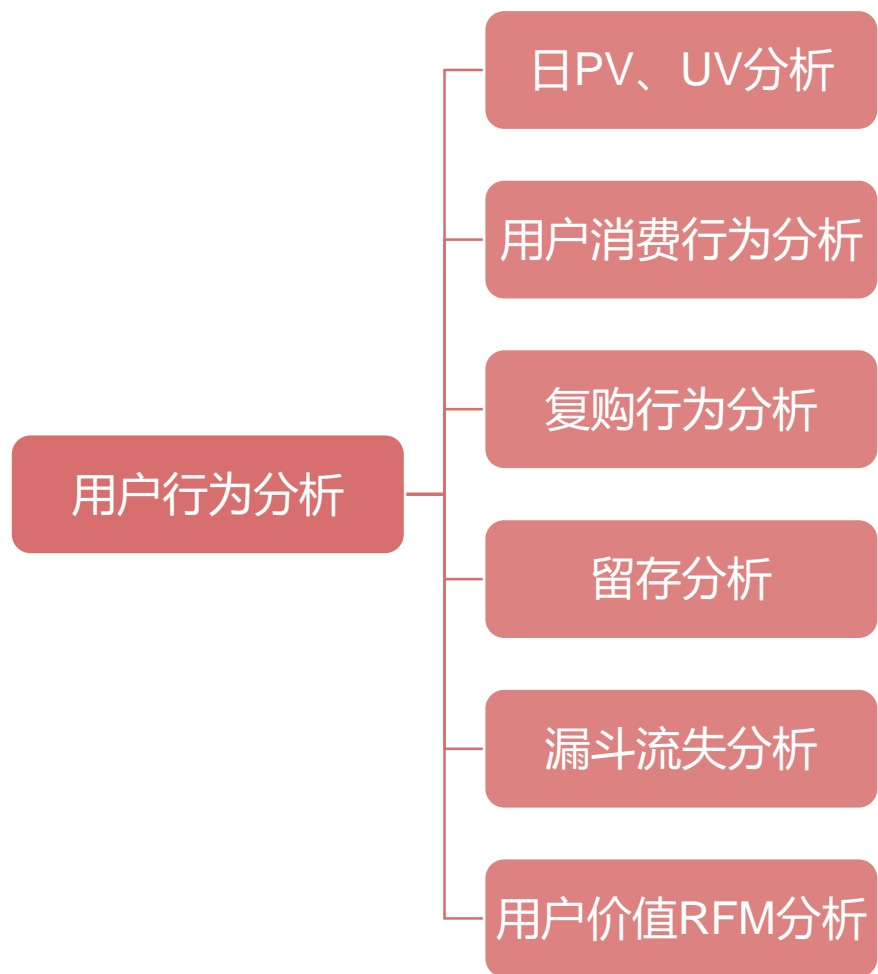
结论

前言

/01



前言



- 针对淘宝app的运营数据，以行业常见指标对用户行为进行分析，来挖掘用户的行为模式。
- 最后针对活跃度，留存，漏斗流失和RFM模型分析结果给出相应的运营建议。

数据来源

/02

数据来源

| | user_id | item_id | behavior_type | user_geohash | item_category | time | hour |
|---|----------|-----------|---------------|--------------|---------------|------------|------|
| 0 | 98047837 | 232431562 | 1 | NaN | 4245 | 2014-12-06 | 2 |
| 1 | 97726136 | 383583590 | 1 | NaN | 5894 | 2014-12-09 | 20 |
| 2 | 98607707 | 64749712 | 1 | NaN | 2883 | 2014-12-18 | 11 |
| 3 | 98662432 | 320593836 | 1 | 96nn52n | 6562 | 2014-12-06 | 10 |
| 4 | 98145908 | 290208520 | 1 | NaN | 13926 | 2014-12-16 | 21 |

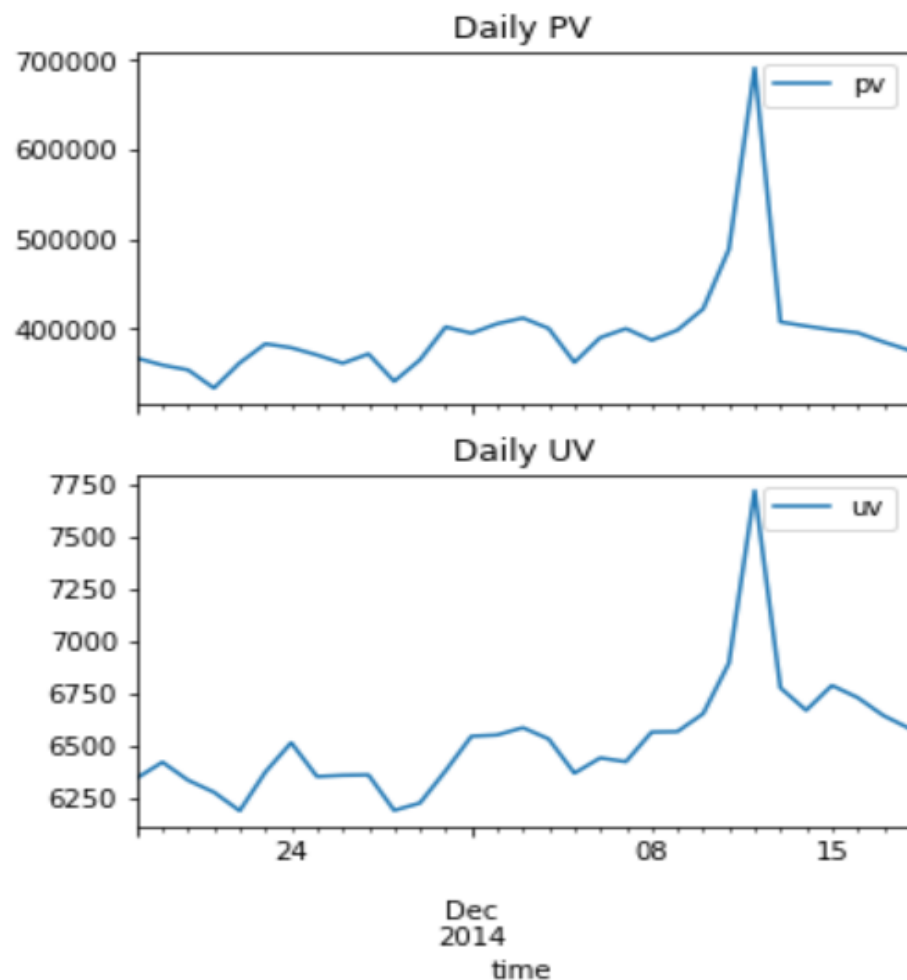
- 数据为淘宝APP **2014年11月18日至2014年12月18日** 的用户行为数据，用户数量10000。
- 该数据包含1048576行，6列，文件名：tianchi_mobile_recommend_train_user.zip，文件大小约0.1GB。
- 数据下载链接：[天池数据集](#)

用户行为分析

/03

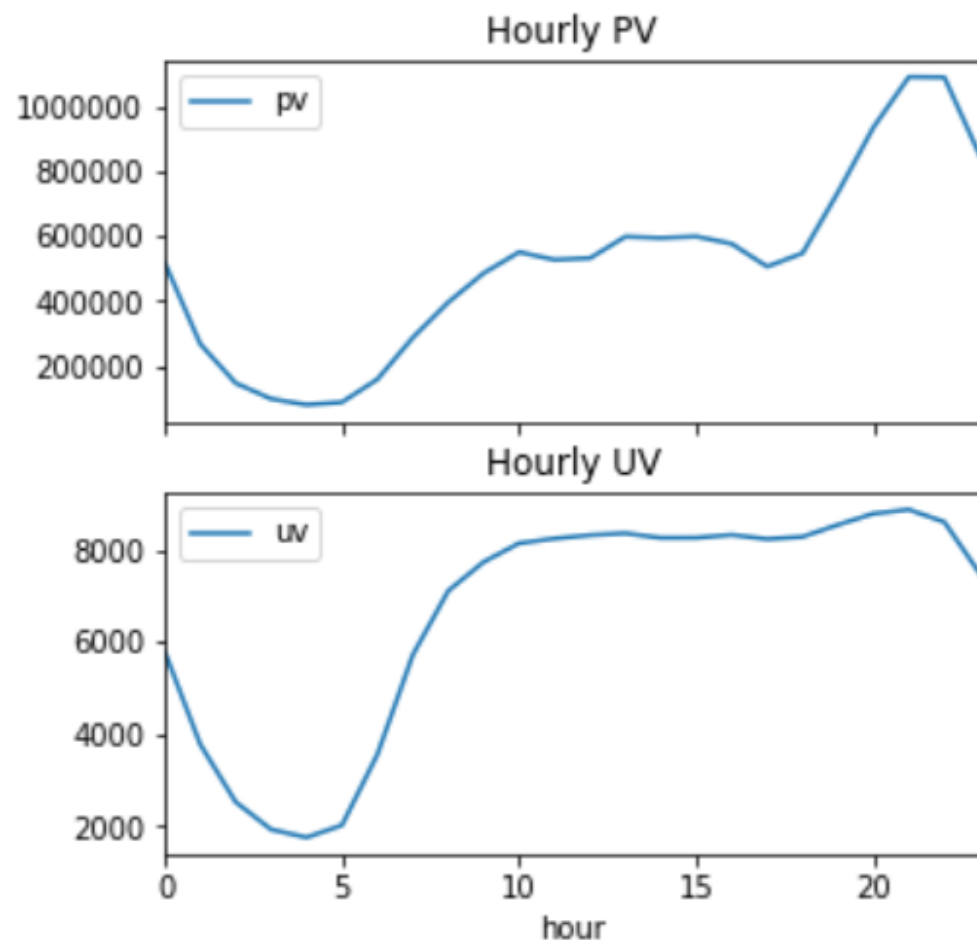


1. 日PV、UV分析



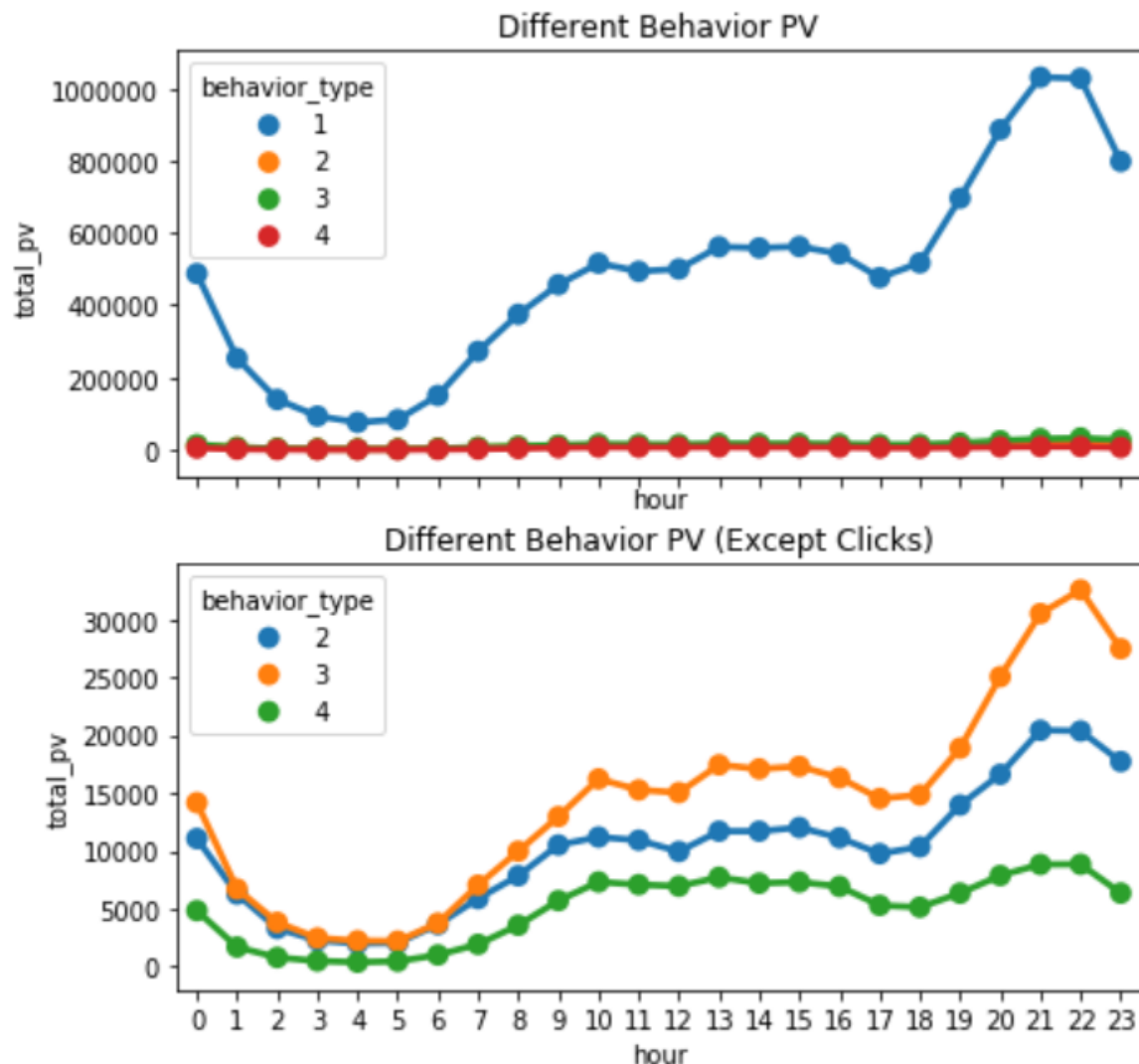
- 可以看出双十二的浏览量最高,并且PV和UV差距很大,但波动情况基本一样;
- 从UV的分布来看, 因为数据总人数为10000, 所以日活率维持在60%以上浮动

1. 日PV、UV分析



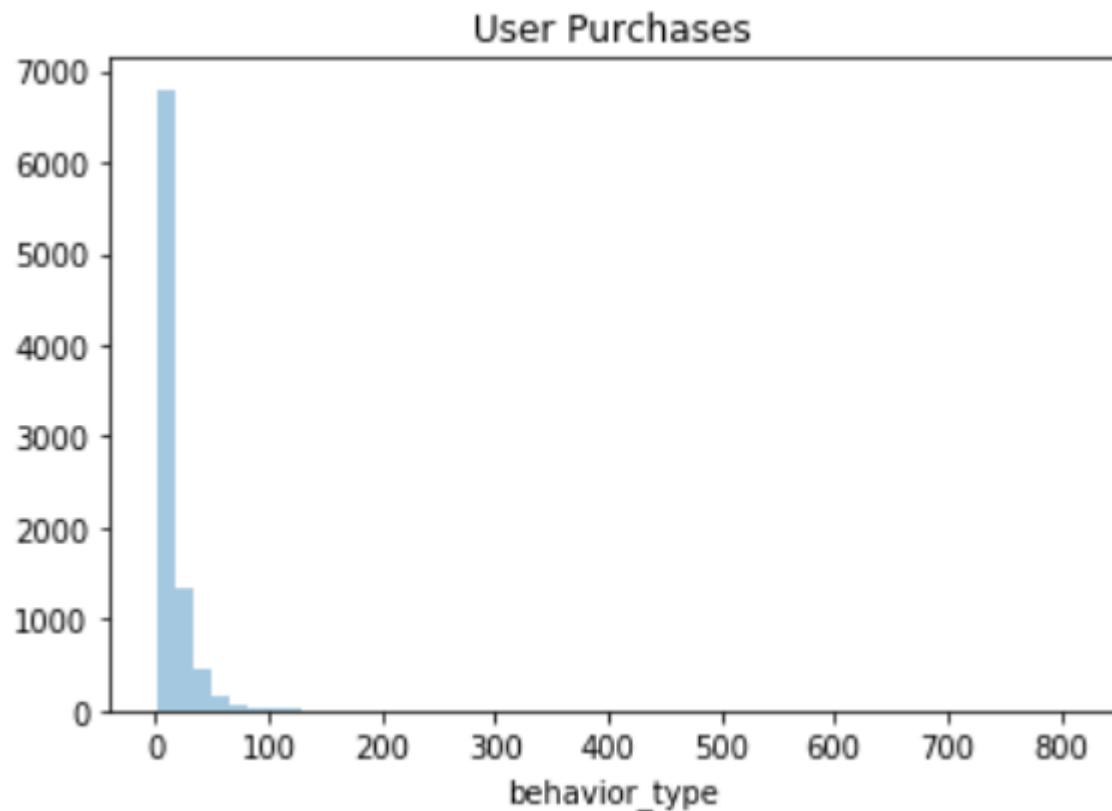
- 可以发现在凌晨时间段PV和UV波动情况相似。
- 但是从晚上大概18点这段时间，UV波动不大，而PV波动巨大，说明晚上是对app操作最频繁的时间段。

1. 日PV、UV分析

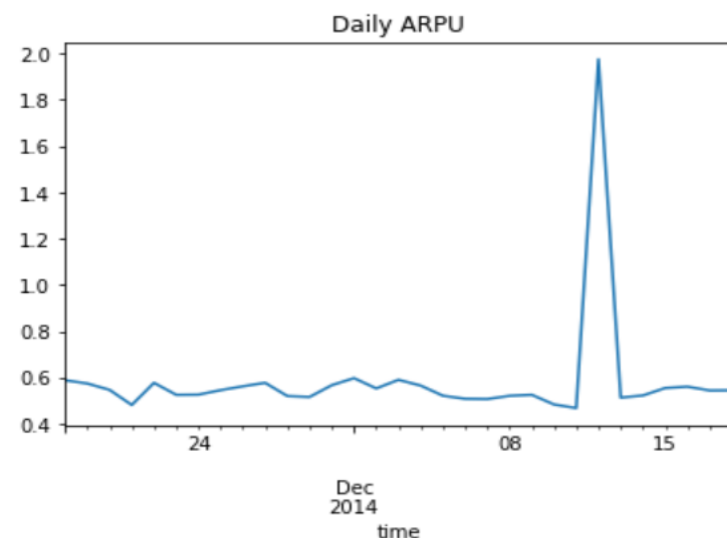
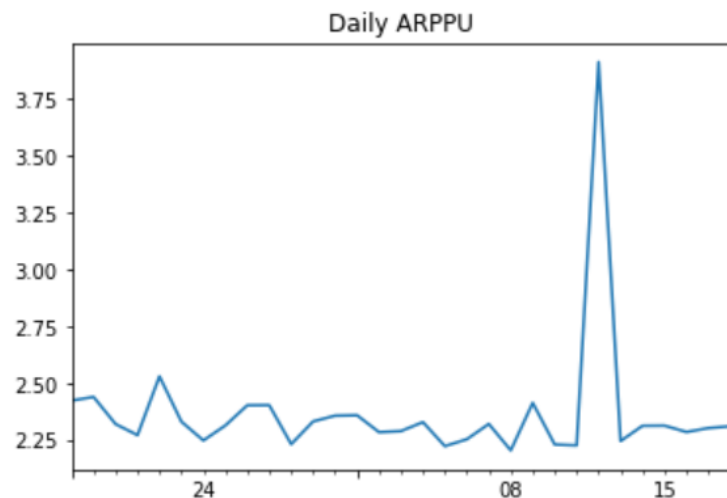


- 1: 点击; 2: 收藏; 3: 加入购物车; 4: 购买
- 可以看出点击情况比其他三类行为多很多, 并且四种行为随着时间的波动情况基本一样。
- 可以了解到用户基本都是在晚上这个时间段有更多的使用行为。
- 同时, 还发现加入购物车的量普遍比收藏量大, 所以在后面漏斗分析可以把加入购物车作为收藏商品的前一步骤。

2. 用户消费行为分析



2. 用户消费行为分析



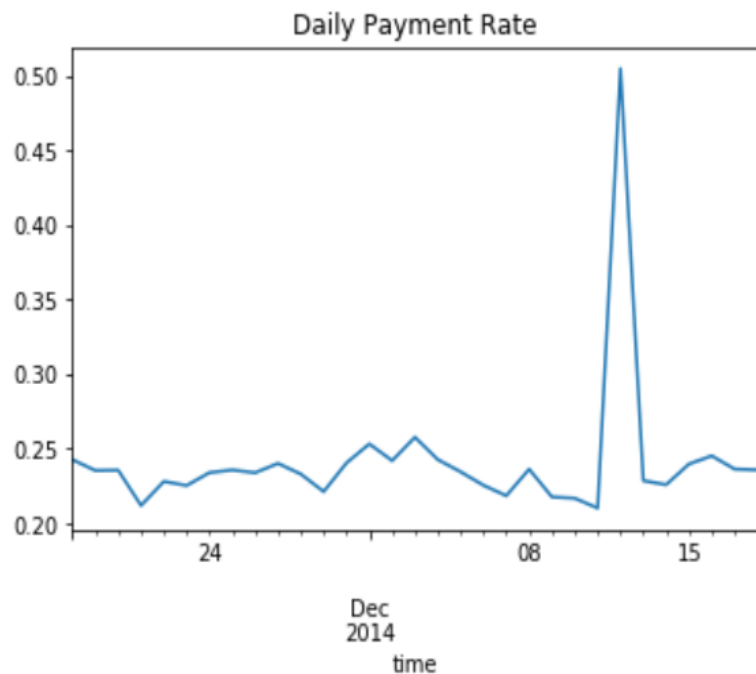
- **日ARPPU情况:**

- 由于数据没有消费金额，因此用消费频率代替金额以做研究。人均消费次数=消费总次数/消费用户人数
- 可以看出基本每天的消费用户平均消费次数在2次以上波动，双十二消费次数比较多。

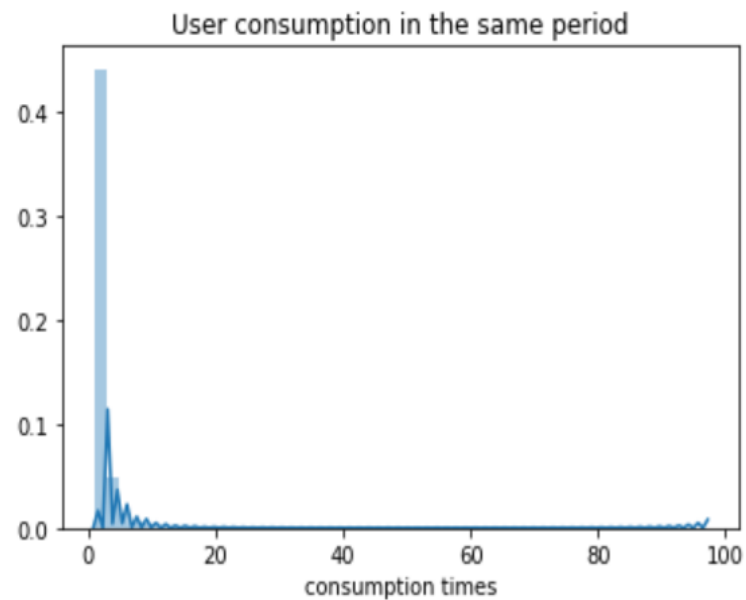
- **日ARPU情况:**

- 活跃用户数平均消费次数=消费总次数/活跃用户人数(每天有操作行为的为活跃)
- 可以看出基本每天活跃用户平均消费次数为0.5左右次。相当于每个人才产生一次消费行为。双十二比较多，达到2次。

2. 用户消费行为分析

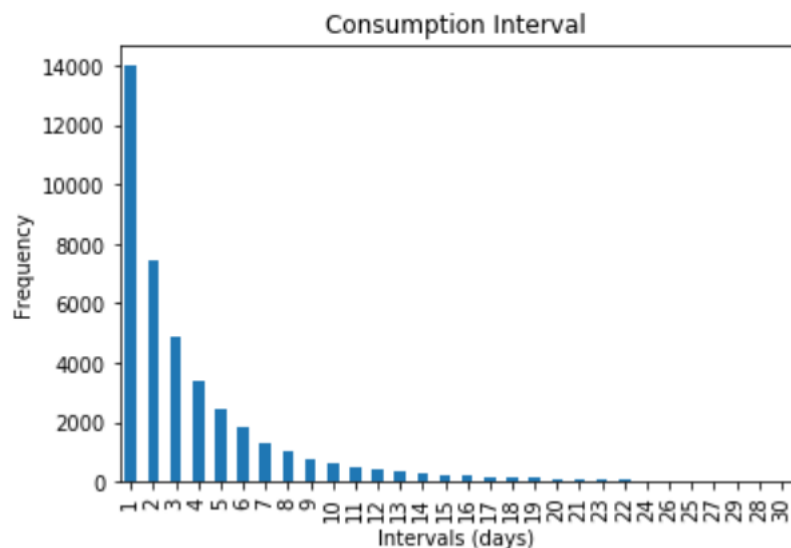


- **日付费率分布情况：**
- 看出每天活跃人群中只有0.2-0.25的活跃人群有消费行为。



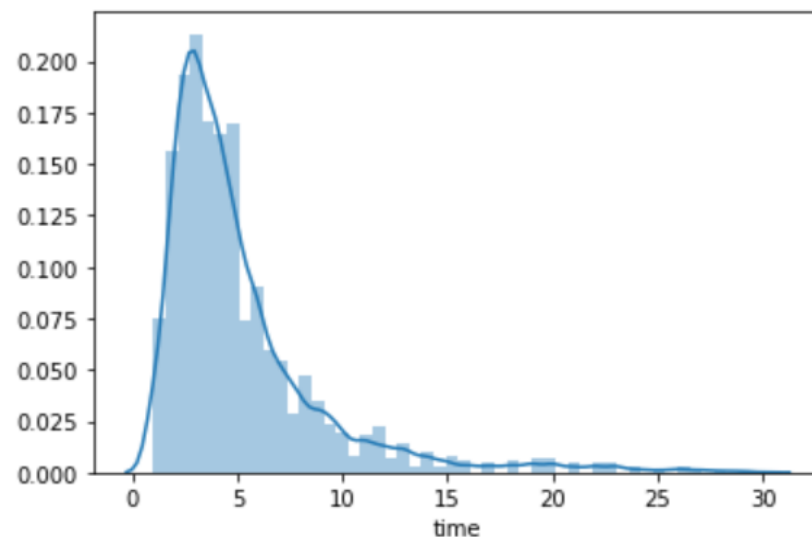
- **同一时间段用户消费次数的分布：**
- 大多数用户消费1次

3. 复购行为分析

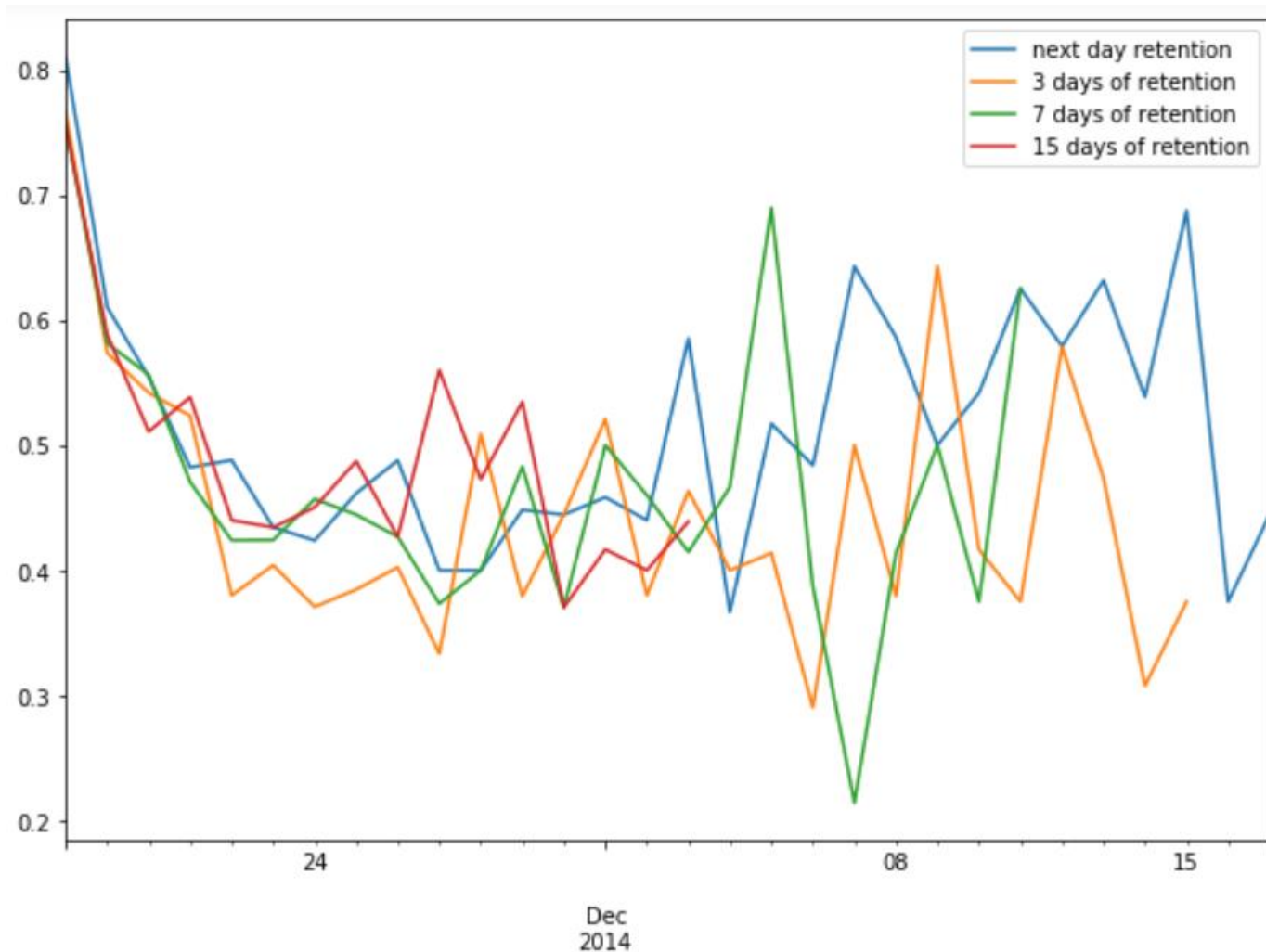


- **消费间隔时间：**
- 发现1-10天间隔呈现急速下降，下降速率不断变缓，到10天后基本很少，说明有复购行为用户发生复购行为间隔基本在1-10天内，很少有人10天后才复购。

- **不同用户平均复购时间：**
- 可以看出整体复购时间随着天数增加呈现递减规律，而对于不同用户的平均复购时间，则呈现一个类正态分布，更多人的平均复购时间为1-5天

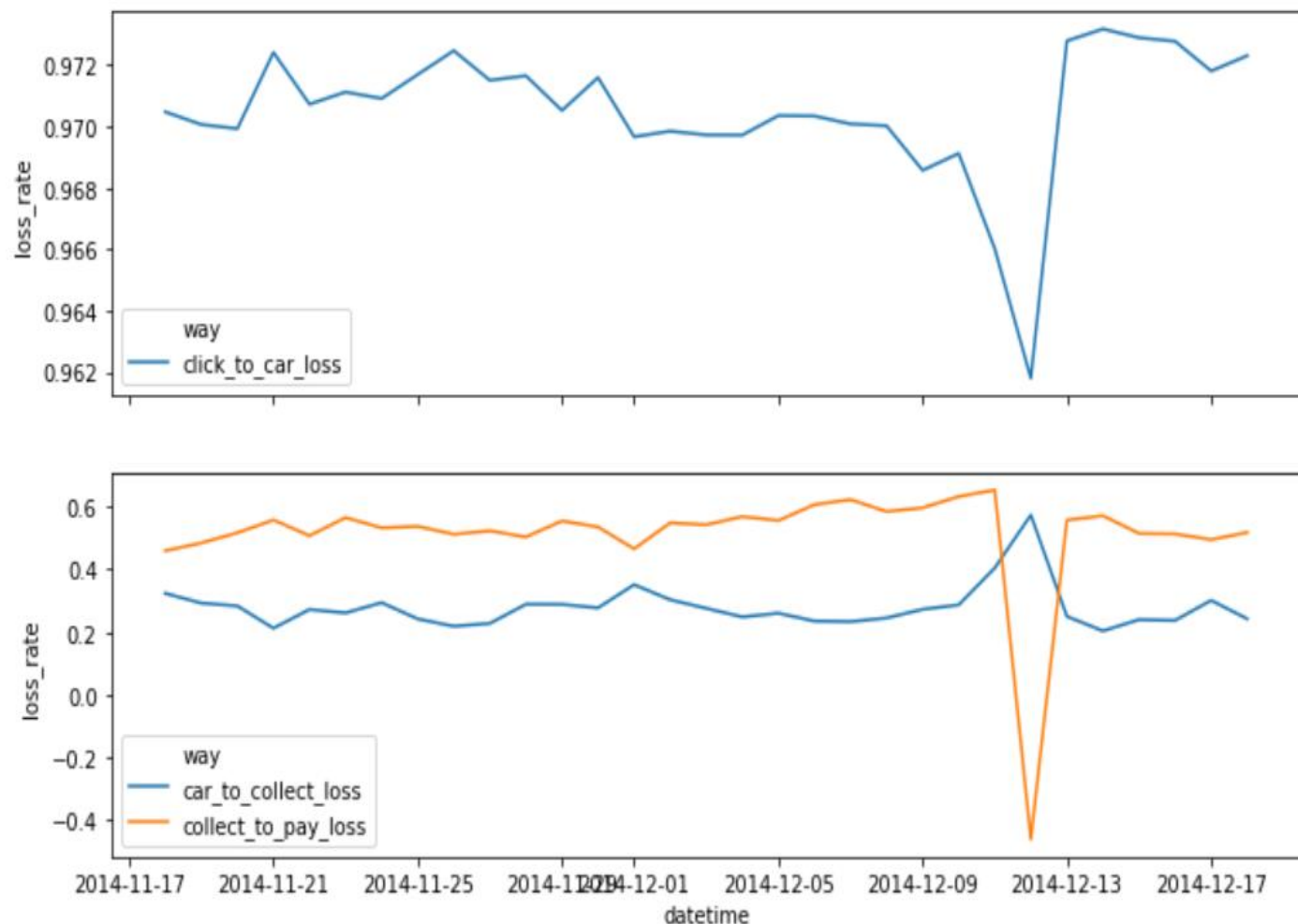


4. 留存分析



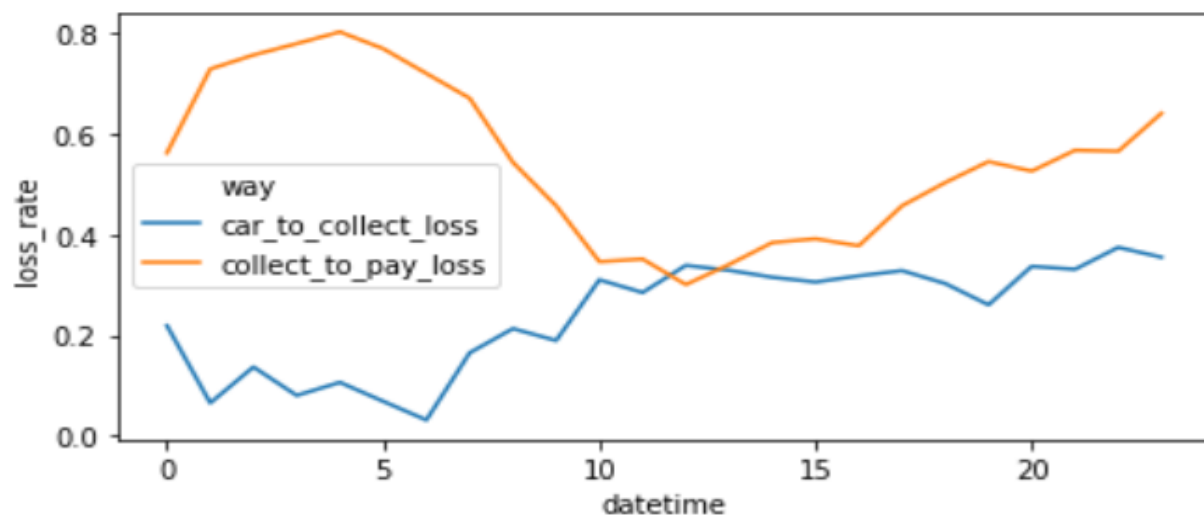
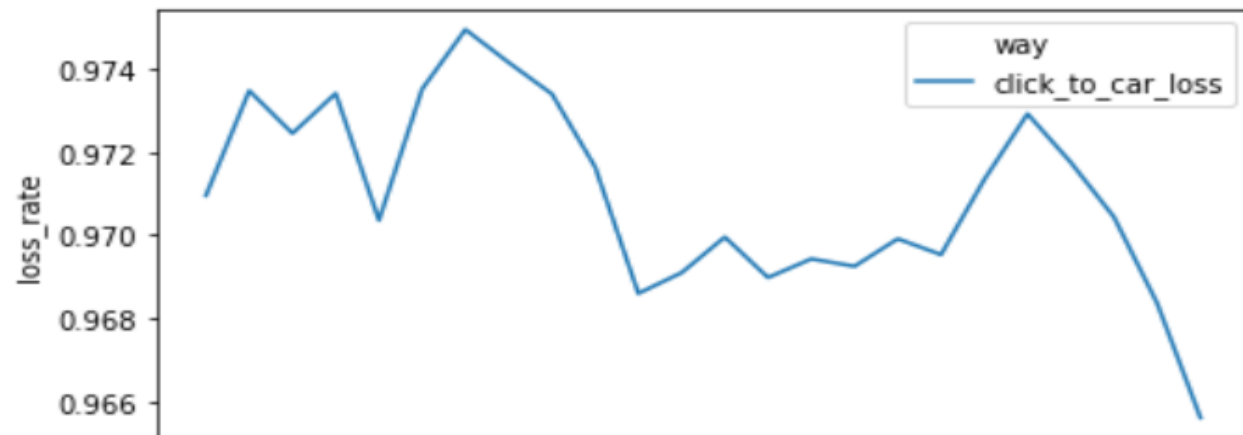
- 由于所选范围一开始时间把第一次登陆作为用户注册时间，这样17号这天会包含大量每天登陆的忠诚客户，所以第一天登陆的用户留存率很高。
- 可以看出留存曲线交叉情况，这说明用户呈现间隔登陆，某日留存并非为用户流失，只是登陆时间不在该天。
- 另外注意12月7号登陆的用户次日，3日，7日留存递减速度很快，这里需要注意，可能这天登陆的用户正在慢慢流失。

5. 漏斗流失分析



- 可以看出点击到加入购物车流失率特别高，在95%以上，而加入购物车到收藏和收藏到支付流失率分别在20%和50%左右，收藏到支付的流失率较高。
- 另外可以看到波动比较稳定，除了双十二，这天点击流失率相对之前偏低，加入购物车到收藏流失变高，而收藏到支付流失则变成负，说明用户都是事先加入购物车然后直接购买的情况大量增加。

5. 漏斗流失分析



- 可以看出凌晨时间段收藏到支付流失率很高，说明这段时间查看商品只为随便看看，产生支付的情况很少，或者预先手残后第二天早上支付。
- 到晚上也是支付流失比较高的，因为这段时间用户有更积极取比较和思考是否要购买。

6. 用户价值度RFM模型分析

| | user_id | recent | freq | recent_value | freq_value | rfm |
|---|-----------|--------|------|--------------|------------|-----|
| 0 | 100001878 | 2 | 36 | 3 | 3 | 33 |
| 1 | 100011562 | 4 | 3 | 3 | 1 | 31 |
| 2 | 100012968 | 2 | 15 | 3 | 3 | 33 |
| 3 | 100014060 | 2 | 24 | 3 | 3 | 33 |
| 4 | 100024529 | 4 | 26 | 3 | 3 | 33 |
| 5 | 100027681 | 6 | 48 | 2 | 3 | 23 |
| 6 | 100035725 | 2 | 4 | 3 | 1 | 31 |
| 7 | 100039408 | 31 | 1 | 1 | 1 | 11 |
| 8 | 100042340 | 2 | 3 | 3 | 1 | 31 |
| 9 | 10004287 | 20 | 14 | 1 | 3 | 13 |

- 由于数据没有金额所以缺少消费金额（m）这一列。通过打分可以了解每位顾客的特性，从而实现**差异化营销**。比如：
- 对于33用户，为重点用户需要关注；
- 对于31这类忠诚度高而购买能力不足的，可以适当给点折扣或捆绑销售来增加用户的购买频率；
- 对于13这类忠诚度不高而购买能力强的，需要关注他们的购物习性做精准化营销。

结论

/04



结论

结论一

通过PV, UV的不同时间维度分析, 可以发现PV, UV随着时间呈现周期性变化, 并且一天中不同时间段用户活跃度也不同, 因此在做商品推广和广告营销尽量选择活跃度比较高的时间段会有更好效果。

结论二

根据复购时间的分析, 可以分析出用户普遍复购时间在3天左右, 因此可以对超过3天未有购买的用户进行适当促销或优惠, 能一定程度提高用户消费积极性。

结论三

通过不同时间漏斗转化分析出流失率比较低的时间段, 在这些时间段进行相应的商品推荐效果更优。

结论四

通过对用户的RFM模型分析, 可以有效对不同用户进行不同群体划分, 不同群体采用不同的个性化营销手段, 能产生更好的效益。



Thanks for watching.

Sudie Chong

THANKS

