

淘宝店铺用户行为分析

一、 分析背景

随着互联网环境逐渐成熟，电商平台的营销策略层出不穷，例如“双十一”、“双十二”就是其中的典型。但是刚刚历经“双十一”为期一周的消费狂欢周后，与之相一致的“双十二”营销是否依然能激发消费动能，促进店铺获客呢？电商又该如何基于“双十二”布局自身的经营策略呢？基于此，本文将基于开放数据集——2014“双十二消费节”期间淘宝消费者行为数据进行分析。

PS：本文档仅用于作者将自身知识用于实践和总结实际业务知识。源码见附件

二、 数据集介绍

本数据集为阿里巴巴天池大赛 2015 年 7 月 1 日赛季竞赛题目所含的消费者行为数据集。包含从 2014 年 11 月 18 日至 2014 年 12 月 18 日之间，10000 名用户共 12256906 条行为数据。数据集每一行表示一条用户行为，其中字段如下：

列字段名称	数据说明+例子
User_id	用户 id, 98047777
Item_id	商品 id, 232325625
Behavior_type	行为: 点击、收藏、加购物车、购买
User_geohash	用户地址, 96nn52
Item_category	商品品类 id, 4245
time	行为发生时间, 2014-11-19 10

三、 明确分析维度

从用户画像、时间分析和用户价值入手，解析用户行为特征变化。聚焦“双十二”活动前后的用户行为特征，以验证“双十二”活动的有效性，基于行为特征为经营赋能。



四、 数据分析

4.1 数据清洗

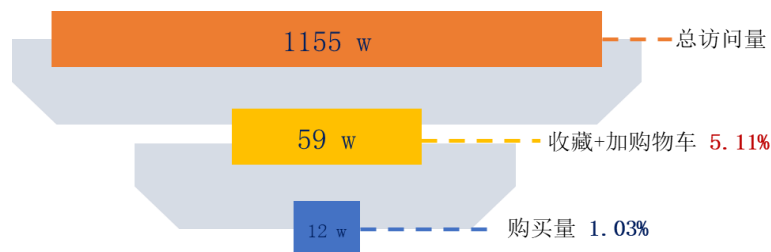
该数据采用 R 进行数据分析和可视化。聚焦用户行为，查看数据全貌，剔除缺失值较多的用户地址信息。由于字段 *time* 既包含日期信息，又包含时间信息，为了简化信息，提高数据可读性，在数据清洗阶段将其分开。同时将 *user_id*, *item_id*, *behavior_type*, *hour* 转化为因子。

```
# 数据清洗
behavior <- read.csv('tianchi_mobile_recommend_train_user.csv') %>% # 读取数据
  select(-user_geohash) %>% # 由于缺失值过多去除此列
# 对时间进行预处理，将时间分为日期，周，时间
mutate(date = as.Date(substring(time, 1, 10), tryFormats = "%Y-%m-%d"),
       week = format(date, format = "%a"),
       hour = substring(time, 12, 13))

# 调整字段格式
behavior$user_id <- as.factor(behavior$user_id)
behavior$item_id <- as.factor(behavior$item_id)
behavior$behavior_type <- factor(behavior$behavior_type, levels = c("1", "2", "3", "4"),
                                labels = c("点击", "收藏", "购物车", "购买"))
behavior$hour <- as.factor(behavior$hour)
```

4.2 用户行为画像

首先，对用户的四种行为的总量及其占比进行对比分析，结果显示，店铺点击总量约为 1155 万次，收藏总量为 24.2 万次，加购物车总量为 34.4 万次，购买总量为 12.0 万次，购买转化率为 1.03%。用户总量为 10000 人，购买用户 8886 人，占比 88.86%。由此可见，店铺用户转化率很高，但用户从点击浏览商品到购买的转化率远低于互联网行业 2% 的转化率，这表明商品虽然精准投放，但存在较长购买潜伏期，存在被竞品替代的潜在风险。



```
# 指标1
# 总点击量（访问量pv）、总收藏量、加购物车量、购买量
table(behavior$behavior_type)
```

```
# 指标3 用户总数(uv)
length(levels(behavior$user_id))
```

```
# 指标4 有购买行为的用户数
behavior[behavior$behavior_type=="购买", ] %>%
  select(user_id) %>%
  unique() %>%
  nrow()
```

对购买转化率低进行析因分析。结果显示，浏览后直接购买的人数为 676 人（0.7%），购买次数/点击量总体均值为 1.6%，与行业水平持平。对加购或收藏后用户转化率进行分析，浏览后加购物车或收藏后购买的人数为 9324 人（0.93%），购物车弃置率为 45.5%，与同行业（70%）相比较好。而访问量到购物车或收藏环节的转化率为 5%。存在较大提升空间。

其次，对被试的日均访问量进行分析并可视化，结果显示（见图 1，word 无法实现交互，可[查看html附件](#)），产品点击次数从 2014 年 11 月 18 日至 2014 年 12 月 18 日之间稳定上升（黑色三角线条），日均访问量在 30-35 万次之间波动，较为稳定。从 12 月 8 日开始至 12 月 12 日呈指数级上升，并在 12 月 12 日当天出现日均访问量的最高峰，随后直线下降至日常水平。由此可见“双十二”营销在当天的效果显著，滞后性较查，因此，对“双十二”的布局可以从 12 月 8 日开始加大营销力度，逐渐提高商品曝光度。

```
# 日均访问量
date_pv <- behavior %>%
  group_by(date, behavior_type) %>%
  summarise(daily_num = n())

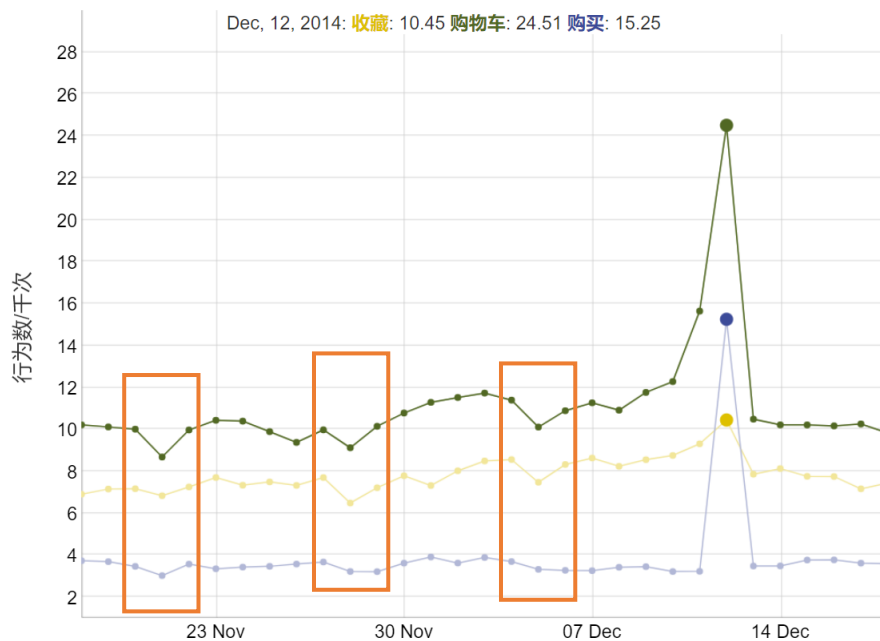
##   date      behavior_type daily_num
##   <date>    <fct>         <int>
## 1 2014-11-18 点击             345855
## 2 2014-11-18 收藏              6904
## 3 2014-11-18 购物车          10212

dygraph(date_dj_final) %>%
  dyAxis("y", label = "点击数/万次", valueRange = c(25, 70)) %>%
  dyAxis("x") %>%
  dySeries('date_dj', strokeWidth = 2, pointShape = "square",
    strokePattern = "dashed", label="点击", color = "#6e7cb9",
    pointSize = 3, drawPoints = TRUE) %>%
  dySeries("date_dj1", axis = 'y2', color = "black orange", label = "累计点击量",
    pointSize = 3, pointShape = "triangle", drawPoints = TRUE) %>%
  dyLegend(show = "follow", width = 175) %>%
  dyOptions(fillGraph = TRUE, fillAlpha = 0.1) %>%
  dyRangeSelector(height = 20, strokeColor = "")
```



图 1 店铺的日均访问量和累计访问量

聚焦收藏、加购物车、购买行为的日均变化，并将其可视化，结果显示（见图 2，交互视图[见html附件](#)），日均加购物车量与日均访问量之间的变化趋势一致且响应紧密；日均收藏量与日均访问量之间的变化响应一般；日均购买量的变化趋势在“双十二”前后差异不大，仅在双十二当天出现激增并达到顶峰。结合 2014 双十二规则，我们发现，在原有活动基础上，仅在双十二当天可以再减去双十二购物券金额购买商品，因此购买行为在仅在该日激增并仅持续一天。由此我们可以得出两个结论：第一，提高活动当日前的曝光度、提高加购物车行为次数有助于促成更多的销售单；第二，结合双十二当日后购买行为的锐减并与之前持平，我们可知，“双十二”官方营销策略为最主要影响因素，且该店铺存在稳定购买需求。在商品营销策略上，可以少使用降价策略，多采用组合售卖策略。



最后，基于用户 id 和行为类型对用户行为进行汇总，将长数据转化为宽数据，得到用户购物行为情况的数据，用于展示每个用户点击、收藏、加购物车和购买次数。

```
# 指标1 用户购物行为分析
behavior %>%
  group_by(user_id, behavior_type) %>%
  summarise(num = n()) -> user_num
user_num <- pivot_wider(user_num, names_from = behavior_type, values_from = num)
user_num[is.na(user_num)] = 0

##   user_id 点击 收藏 购物车 购买
##   <fct>   <int> <int>  <int> <int>
## 1 4913    1658   49     29     6
## 2 6118     112    4      0     1
## 3 7528     183    1     24     6
```

4.3 基于时间的消费者行为分析

面临“双十二”营销规则日趋复杂，甚至每天、每个时段的规则都有所不同，因此，在时间序列上，分别以一周和一天为刻度对用户行为的时间变化进行分析。

首先，在星期维度上，我们发现将“双十二”当天纳入考量时，收藏加购日均次数在周二、周三、周四均存在明显上升，并在周五显著下降至常规水平；另外购买行为在周五激增，而在一周内的其余时间保持稳定。因此将双十二当天剔除后，以日均行为量为指标重新可视化。结果显示购买行为在周一至周五均持平，而收藏加购依然存在上升趋势，这表明双十二对店铺引流的作用明显；另外，结合日均变化图（图 1），我们发现，该产品每至周五在访问量、收藏量、加购物车量上都会下降。

```
# 删去点击率计算周变化，发现周五的购买存在异常
ggplot(data = week_num[week_num$behavior_type != "点击",],
       aes(x=week,y=数量/1000,color=behavior_type,group = behavior_type))+
  geom_line(size=1.2)+
  geom_point(size=2)+
  theme_ipsum()+
  labs(x="",y='次数/k次')+
  scale_color_manual(values = met.brewer("VanGogh2",4)[2:4])+
  scale_y_continuous(limits = c(0,60))+
  theme(axis.text.x = element_text(size = 8),legend.title = element_blank(),
        legend.position = c(0.87,0.9))

ggplot(data = week_num1[week_num1$behavior_type != "点击",],
       aes(x=week,y=数量/1000,color=behavior_type,group = behavior_type))+
  geom_line(size=1.2)+
  geom_point(size=2)+
  theme_ipsum()+
  labs(x="",y='次数/k次')+
  scale_color_manual(values = met.brewer("VanGogh2",4)[2:4])+
  scale_y_continuous(limits = c(0,15))+
  theme(axis.text.x = element_text(size = 8),legend.title = element_blank(),
        legend.position = c(0.87,0.9))
```

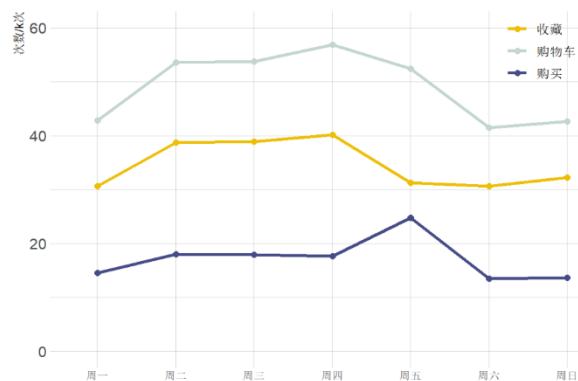


图 2-1 未剔除 12.12 当天

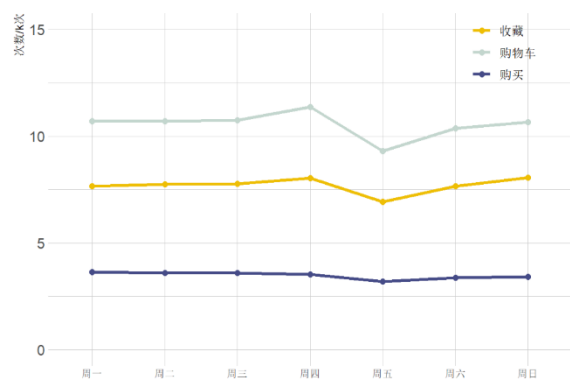
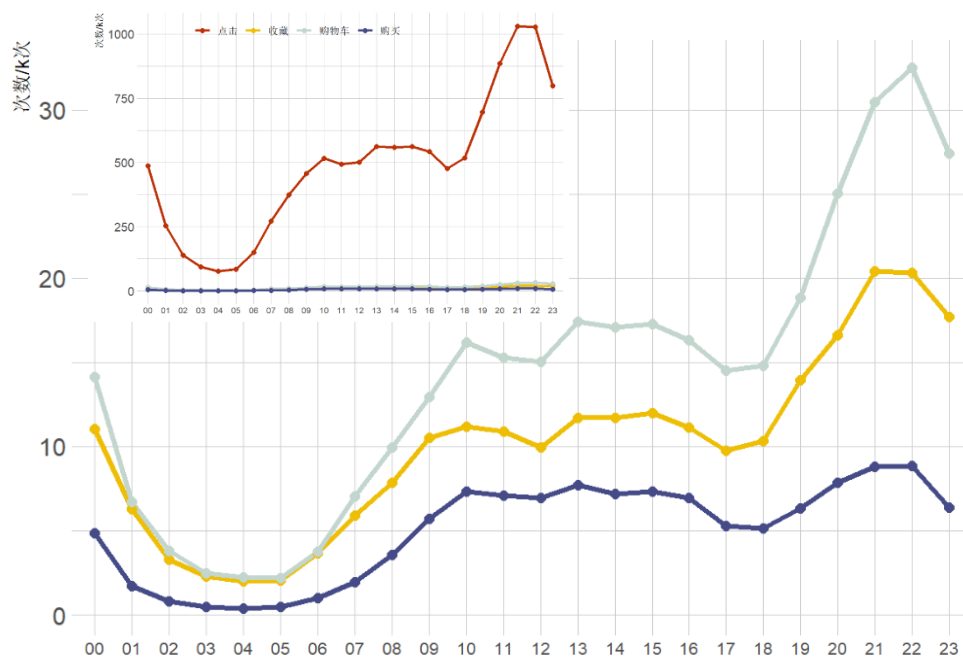
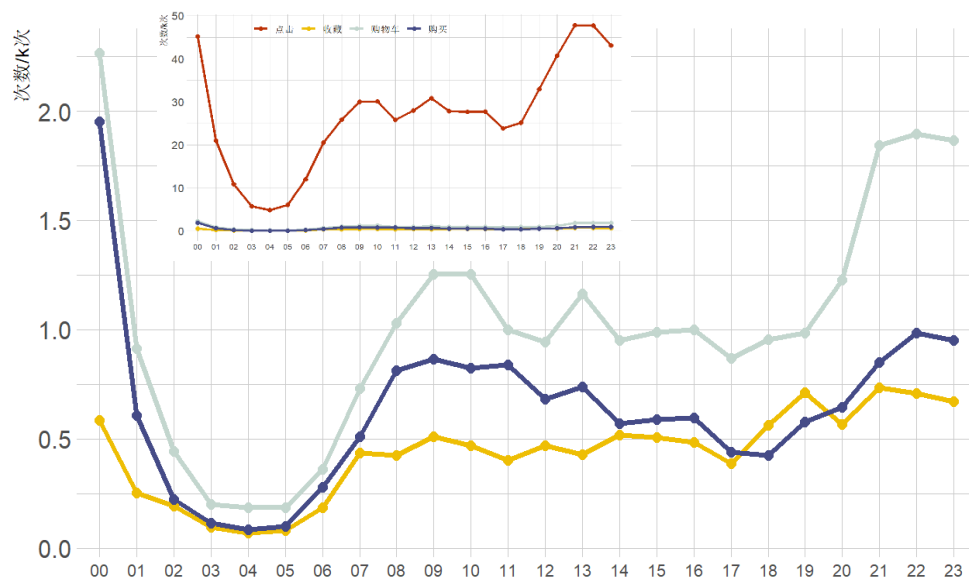


图 2-2 剔除 12.12 当天

用户在不同时间段内的表现如何呢？以天为刻度的可视化结果显示，点击量、收藏量、加购量和购买量的变化趋势基本一致。在凌晨 1 点至 6 点用户睡眠时间段内较低为一日最低点，但在随后逐渐上身，白天 9: 00-17: 00 内基本持平，在晚饭时间 17: 00-19: 00 间稍稍下降，而在晚间 19: 00-23: 00 间访问量、收藏量均显著上升，表明用户在晚上又更多的时间浏览商品，而购买行为却没有显著增加，这表明店铺商品性质必然是存在稳定需求的商品，冲动性购买较少。



对“双十二”前后一天及“双十二”当天进行探索性分析，结果显示，仅在“双十二”当天各指标变化曲线存在明显不同。由图可知，在“双十二”当日，0点-1点、21点-23间流量达到最高峰，但购买行为并未在21点至23点间存在峰值。因此，在制定营销策略时，建议活动前力度大，活动当日后及时调整。



4.4 用户价值分析

运用 RFM 法对发生购买行为的用户价值进行分类，RFM 分别是指 Recency（最近消费间隔）、Frequency（消费频率），Monetary（消费金额）。R 用于描绘用户敏感性，一般认为消费间隔越短，对产品的感受越真切；F 用于描绘用户的忠诚度；根据二八法则，M 用于描绘用户的价值。由于该文档不包含消费额字段，因此不对 M 进行潜在分析。

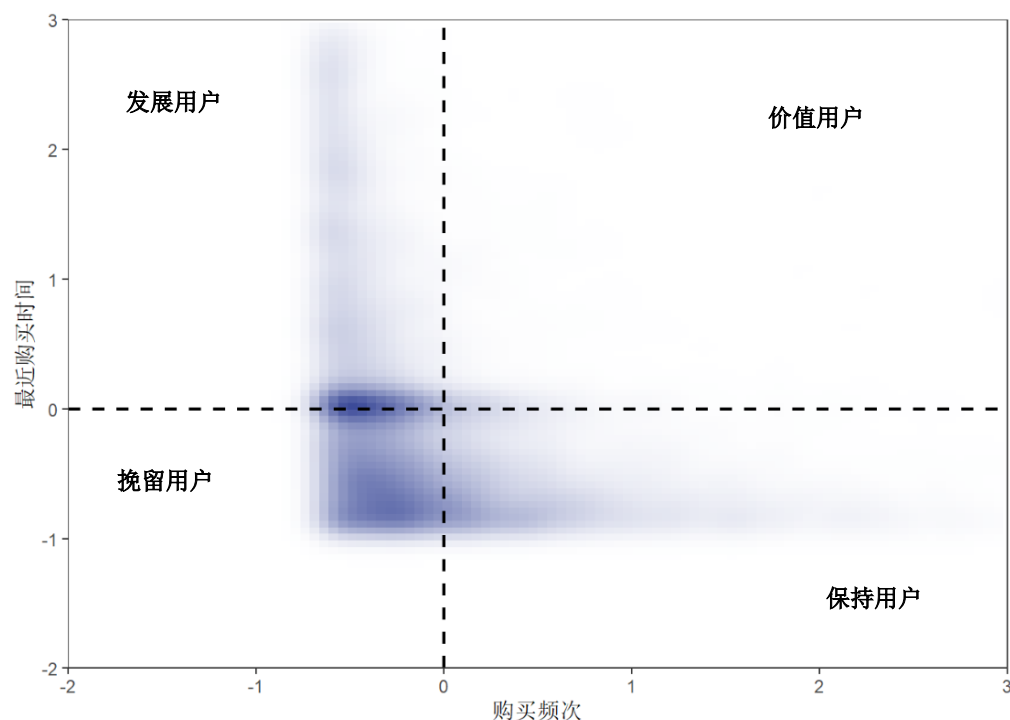
首先，计算用户的 R 和 F 值，并对其评级评分。其中消费频率为前 25%（17 次）评为

4分，前25%~50%（8~17次）评为3分，前75%~50%（3~8次）评为2分，最后25%（1~3次）评为1分。

```
RFM_1 <- behavior[behavior$behavior_type == "购买",] %>%
  group_by(user_id) %>%
  summarise(R = difftime("2014-12-19", max(date), units="days"), Fre = n()) %>%
  arrange(R) %>%
  mutate(rrank = row_number(),
         fscore=case_when(Fre > quantile(Fre)[4] ~ 4,
                          Fre > quantile(Fre)[3] & Fre <= quantile(Fre)[4] ~ 3,
                          Fre > quantile(Fre)[2] & Fre <= quantile(Fre)[3] ~ 2,
                          Fre <= quantile(Fre)[2] ~ 1))
```

根据 RFM 模型的分类，将用户分为价值用户、发展用户、保持用户和挽留用户这四种。将得分中心化后以密度图进行描绘。从图中可以看出，用户主要集中在挽留用户上，价值用户较少。对于挽留客户和保持客服，最近一次购买时间较远，存在潜在流失风险，需要主动电联客服，调查问题，提高复购率。对于发展用户，需要想办法提高用户的购买频率，如推送优惠信息等。

```
ggplot(RFM_1, aes(x=scale(Fre), y=scale(R)))+
  stat_density2d(aes(alpha=..density..), fill = "#3c4b99",
                geom="tile", contour=FALSE, show.legend = FALSE)+
  geom_vline(aes(xintercept=0), linetype=2, size=0.8)+
  geom_hline(aes(yintercept=0), linetype=2, size=0.8)+
  theme_bw()+
  scale_alpha_continuous(range = c(0, 1))+
  scale_x_continuous(limits = c(-2, 3), expand = c(0, 0))+
  scale_y_continuous(limits = c(-2, 3), expand = c(0, 0))+
  theme(panel.grid = element_blank())+
  labs(x="购买频次", y="最近购买时间")
```



五、 结论

- 1. 从用户行为总体情况来看**，“双十二”购物节在引流上的效果是明显的，但是对购物的影响仅在购物当天是显著的，其余时间购买行为与日常一致。这可能是由于商品性质决定的——商品是具有稳定需求的通用必需品。另外，点击量和购买量之间变化趋势随时间相一致，因此，对于“双十二”的营销布局可以通过增加“双十二”当天的曝光度来增加购买行为。
- 2. 从购买变化趋势来看**，“双十二”后并未出现购买量下浮，而是恢复至之前水平。这表明“双十二”营销仅仅在购买上是成功的，但店铺策略并未借助“双十二”的余温，提升店铺后续购买量。在策略上在“双十二”后也要注重用户留存，提高转化率；另外，不宜采取降价优惠的方式保持留存。
- 3. 从转化率上来看**，点击转化率（1%）远低于行业标准，从主要因素上来看，原因在于访问到购物车或收藏环节的转化率过低（5%）。建议增加产品亮点以提高顾客加购行为。
- 4. 从访问量和购买量的时间分布来看**，用户在非双十二当日，晚上 20 点-23 点间的点击量最大，而在双十二当日，凌晨 0 点-1 点的点击量和购买量最大。因此，在平时可以在晚间加大营销力度，增加购买行为。
- 5. 从客户价值分类上来看**，由于产品自身特性可能是存在周期性的必需品，从 11 月 18 日-12 月 18 日之间的用户也都以挽留用户和保持客户为主，价值用户很少。因此，在营销策略上，要以增加用户忠实度、避免潜在竞品风险为主。建议主动电话联系用户，周期性推广产品情况、建立营销社群等方式增加店铺产品曝光。