



专注于商业智能BI和大数据的垂直社区平台

# 收入分析实战

谢佳标 ( Daniel.xie )

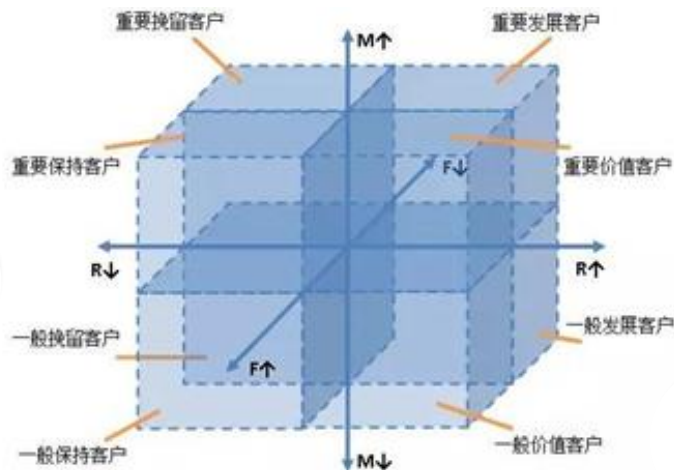
# RFM模型原理

- 通过某个客户近期的购买行为、购买频率和消费金额三个指标来描述客户的价值状态。

## 消费金额 ( Monetary )

M

- 单位期间内的消费总金额或平均消费额。
- 根据M值越大越好的标准，总额越大越有可能再次购买



R

## 最近一次消费 ( Recency )

- 客户最近一次的购买时间是什么时候。最近一次消费时间越近的顾客是最有可能对提供的商品或是服务也最有反应的群体。
- 根据R值越来越小的标准，间隔时间越短越有可能再次购买

F

## 消费频率 ( Frequency )

- 单位时间内的消费次数。最常购买的顾客，也是满意度最高的顾客。
- 根据F值越多越好的标准，单位时间内消费次数越多越有可能再次购买

# RFM模型用户群特征

RFM类型	用户类型	特征	用户重要性	应对策略
011	重要保持客户	这类客户购买频繁、购买金额大且最近一次交易时间间隔短	1	客户实际贡献的价值很高，且具有很高的潜在价值，是公司的优质客户群，公司利润的主要贡献者，因此可视为公司的 important 保持客户，继续维持与这类客户的关系是公司利润的重要保障。
001	重要发展客户	这类客户最近一次交易时间间隔短、购买金额大，购买频率较低	2	对公司的利润贡献不及“011”型客户；但是，这类客户具有很高的潜在价值，如果公司分析、了解、满足他们的需求，利用针对性的营销手段吸引他们，提高购买频率，将给公司带来更多利润，因此这类客户可视为公司重要的发展客户
010	一般重要客户	这类客户最近一次交易时间间隔短，购买频率高，属于活跃客户，但累计购买交易金额较少，公司利润也较少。	4	这类客户有可能购买力有限；也可能购买力强，但对公司的一些产品不感兴趣。加大这类客户的营销投入存在一定的失败风险，但适当维持与这些客户的关系又能使公司获得一定利润。因此，这类客户属于公司的一般重要客户。
000	(重要发展客户 or 低价值用户)	这类客户最近一次交易时间间隔短，但购买频率和购买量都较低，无法立即给公司带来丰厚利润；	2 or 6	如果他们属于新客户，那么是公司扩大客户量和市场份额的重要客户源，属于重要发展客户；如果属于老客户，则是低价值客户。
111	重要挽留客户	这类客户与公司的购买频率很高、购买量也很多，但长时间没有与公司交易，存在流失风险。	3	对这类客户，公司应尽量挽留，通过营销手段提高客户忠诚度，让客户继续留在公司。因此可视为公司的重要挽留客户，是公司利润的潜在来源之一。
110	一般客户	这类客户购买频率较高，但长时间没有与公司交易，而且购买量很低，公司已很难从他们身上获取更多利润，因此，只能看作公司的一般客户	5	一般对待
100	低价值客户	从购买频率，购买量及购买近度三方面分析，这类客户都属于“劣质”客户	6	公司没有任何必要继续维持与他们的关系，属于公司的无价值客户。
101	一般客户	虽然购买金额较大，但购买频率低和购买间隔长，不是公司的忠诚客户，他们与公司的交易存在偶然性，可视为公司的一般客户。	5	一般对待

# 业务案例：付费用户RFM模型研究

- 有一款游戏最近一周的付费类数据，包括本周最后一次付费日期、付费次数、付费金额等。现在将数据导入R中，进行数据处理，选择合适的划分方法。

# 导入RFM数据

```
rfm_data <- read.csv("RFM_data.csv")
```

# 将最后付费日期转换成日期格式

```
rfm_data$last_date <- paste(substr(rfm_data$last_date,1,4),substr(rfm_data$last_date,5,6),  
                             substr(rfm_data$last_date,7,8),sep="/")
```

```
rfm_data$last_date <- as.Date(rfm_data$last_date,"%Y/%m/%d")
```

# 增加距离统计日的相隔天数

```
rfm_data$time_internal <- max(rfm_data$last_date)-rfm_data$last_date
```

# 查看数据的前六行

```
head(rfm_data)
```

# 查看不同指标的10%数据情况

```
quantile(rfm_data$time_internal,probs = seq(0,1,0.1))
```

```
quantile(rfm_data$pay_mnt,probs = seq(0,1,0.1))
```

```
quantile(rfm_data$pay_cnt,probs = seq(0,1,0.1))
```

# 业务案例：付费用户RFM模型研究

- 从百分比结果可以看出，不同指标的数据存在明显左偏现象，用传统的按照平均数进行分箱的方法不适合此数据。我们选择按照百分比的实际结果作为划分标准，确定各个维度基准值。其中R的基准值取70%分位数0天，F的基准值取60%分位数的1次，M的基准值取50%分位数的6元。由此产生新的三个衍生变量：tagR、tagF、tagM。当值大于基准值时，tagR为0，tagF、tagM为1，当值小于等于基准值时，tagR为1，tagF、tagM为0。

# 建立新的衍生指标

```
rfm_data$tagR <- ifelse(rfm_data$time_internal <= 0,1,0)
```

```
rfm_data$tagF <- ifelse(rfm_data$pay_cnt <= 1,0,1)
```

```
rfm_data$tagM <- ifelse(rfm_data$pay_mnt <= 6,0,1)
```

```
head(rfm_data)
```

# 业务案例：付费用户RFM模型研究

- 这样每个用户就有3个标识，分别为tagR、tagF、tagM，每个标识有0和1两种取值，因此用户类型可以分为 $2^3=8$ 种。根据前面的RFM分类给每位用户打上相应的类型标签，统计不同RFM类型的用户人数和付费金额，并绘制柱状图。

```
# 给用户打上类型标签
rfm_data$type <- "一般挽留用户"
rfm_data[rfm_data$tagR==1 & rfm_data$tagF==1 & rfm_data$tagM==1,'type'] <- "重要保持客户"
rfm_data[rfm_data$tagR==1 & rfm_data$tagF==0 & rfm_data$tagM==1,'type'] <- "重要发展客户"
rfm_data[rfm_data$tagR==1 & rfm_data$tagF==1 & rfm_data$tagM==0,'type'] <- "一般保持客户"
rfm_data[rfm_data$tagR==1 & rfm_data$tagF==0 & rfm_data$tagM==0,'type'] <- "一般客户"
rfm_data[rfm_data$tagR==0 & rfm_data$tagF==1 & rfm_data$tagM==1,'type'] <- "重要挽留客户"
rfm_data[rfm_data$tagR==0 & rfm_data$tagF==1 & rfm_data$tagM==0,'type'] <- "一般客户"
rfm_data[rfm_data$tagR==0 & rfm_data$tagF==0 & rfm_data$tagM==0,'type'] <- "低价值客户"
head(rfm_data)
# 统计不同用户群的人数
library(sqldf)
rfm_data_sum <- sqldf("select type,count(distinct player_id) as usr_cnt,
                        sum(pay_mnt) as pay_mnt_sum from rfm_data group by type")
# 按照付费金额进行降序排列
rfm_data_sum <- rfm_data_sum[order(rfm_data_sum$pay_mnt_sum,decreasing = T),]
rfm_data_sum$usr_cnt_rate <- paste0(round(rfm_data_sum$usr_cnt*100/
                                         sum(rfm_data_sum$usr_cnt),2),"%")
rfm_data_sum$pay_mnt_sum_rate <- paste0(round(rfm_data_sum$pay_mnt_sum*100/
                                              sum(rfm_data_sum$pay_mnt_sum),2),"%")
knitr::kable(rfm_data_sum)
```

	type	usr_cnt	pay_mnt_sum	usr_cnt_rate	pay_mnt_sum_rate
5	重要保持客户	1098	143775.00	29.03%	81.24%
6	重要发展客户	378	16717.00	9.99%	9.45%
7	重要挽留客户	231	7567.00	6.11%	4.28%
3	一般客户	1313	4715.03	34.72%	2.66%
4	一般挽留用户	41	2072.00	1.08%	1.17%
1	低价值客户	601	1876.00	15.89%	1.06%
2	一般保持客户	120	243.00	3.17%	0.14%



# 传统RFM模型的不足及改进

## 传统RFM

- 用户评分过于简单
- 用户行为变化无法追踪
- 用户潜在行为难以捕捉
- RFM的权重度量困难
- 更多维度描述客户（最大付费额，最小付费额，最大付费频次等等）——主成分分析（用户属性归一化+属性降维）
- 用户忠诚度评价——主成分分析+权重划分
- 属性权重=各属性成分之和/主成分方差贡献率之和

## 改进优势

- 用户评分不是离散型值，而是连续的忠诚度评分。
- 用户分类通过实际数据（聚类）来决定，而不是简单经验值分类。
- 运营更加详细的了解每个属性的权重值，并且可以细致化分析每个属性对用户忠诚度的影响
- 支持灵活可扩展的运营策略。
- 选择何种策略来挽留用户。

# 业务案例：游戏经济健康度与用户关系分析

- 背景及数据介绍：
  - 某款棋牌游戏经济系统包含了游戏币发放总量、纯消耗、玩家身上携带与保险箱结余等货币量信息。但是这些数据都没能很好地利用起来，运营人员不清楚游戏整个经济系统现状，无法判断目前处于游戏币发放过少还是过剩状态。本次研究希望通过货币类指标与用户类指标相结合，利用用户行为数据研究关键货币量指标，找到阈值判断当前外网货币库存量的健康状况。

指标类型	指标名称	指标类型	指标名称
货币量指标	沉睡保险箱结余	用户类指标	日活跃用户
	沉睡身上携带		月活跃用户数
	活跃保险箱余额		日活跃/月活跃
	活跃身上携带		登陆用户数
	游戏币发放		日新增用户
	纯消耗		累增用户
运营类指标	日充值总额		日付费用户数
	日 ARPPU		首付用户数
	在线均值		付费次数
	在玩均值		玩牌人数
	最高在线		没玩牌人数
	在玩峰值		破产用户数



# 业务案例：游戏经济健康度与用户关系分析

- 数据探索性分析：相关性分析

- 利用相关性分析研究货币量数据与用户属性间的关系，发现对沉淀货币量的变化影响大（通过相关系数体现）的用户行为指标，通过业务运营方式调整对应的用户行为指标，达到减少沉淀货币量的效果；发现对活跃货币量的变化影响大的用户行为指标，通过运营手段提高活跃货币量。

# 数据探索分析

```
w <- read.csv("经济系统数据.csv",T) # 导入数据
```

# 1. 相关性分析

# 研究货币量与活跃用户、新增用户间的关系

```
w <- w[,-1] # 删除日期
```

```
w1 <- w[,c(1:10,16:17)] #选择货币量类与活跃用户类、新增用户类等字段的数据
```

```
(r1 <- round(cor(w1[,1:6],w1[,7:12]),2)) # 计算相关系数
```

```
library(corrplot)
```

```
corrplot(r1,method = "ellipse") # 绘制相关系数图
```

# 研究货币量与付费用户间的关系

```
w2 <- w[,c(1:6,11:15)] # 选择货币量与付费用户的字段数据
```

```
(r2 <- round(cor(w2[,1:6],w2[,7:11]),2)) # 计算相关系数
```

```
corrplot(r2,method = "ellipse") # 绘制相关系数图
```

# 货币量与用户玩牌数据间的关系研究

```
w3 <- w[,c(1:6,18:26)] # 选择货币量与用户玩牌数据户的字段数据
```

```
(r3 <- round(cor(w3[,1:6],w3[,7:11]),2)) # 计算相关系数
```

```
corrplot(r3,method = "ellipse") # 绘制相关系数图
```

# 业务案例：游戏经济健康度与用户关系分析

- 数据探索性分析：降维及聚类分析

- 首先利用主成分分析进行降维处理。利用 scale函数对数据做z-score处理，princomp函数建立主成分分析模型。然后利用聚类分析模型对加工后的数据根据欧氏距离进行层次聚类，从聚类结果中发现与货币量聚在一起的关键用户行为指标。利用dist函数计算欧氏距离矩阵，并利用hclust函数建立层次聚类模型，最后通过plot函数绘制聚类树形图。

```
w.sc<-scale(w) # 将数据标准化
w.pr<-princomp(w.sc) # 建立主成分分析模型
summary(w.pr) # 提取主成分信息
screeplot(w.pr,type="line") # 绘制主成分的碎石图
abline(h=1)
x<-data.frame(w.pr$loadings[,1:5])
dist.e=dist(x,method="euclidean") #计算其欧氏距离矩阵
model1=hclust(dist.e,method="complete") #建立系统聚类模型
plot(model1,main='聚类树图') #画聚类树图
```

# 业务案例：游戏经济健康度与用户关系分析

## • 模型构建：

- DAU/MAU：该指数，用于衡量用户的黏性，留存率，游戏收益情况。这在社交游戏中使用率非常高，可以认定为用户活跃度指数，理论上可以接受的风险值是0.2，也就是说当值低于0.2时，游戏的整体服务进入一个衰退的阶段。利用DAU/MAU=0.2临界点作为风险值依据，研究衡量外网货币库存量的健康指标。通过货币健康度来对游戏币发放进行策略指导。

# 构建衍生指标

```
cadr <- round(w$纯消耗/w$游戏币发放,2)
```

```
smpc <- (w$沉睡保险箱结余+w$沉睡身上携带)/(w$累增用户-w$月活跃用户数)
```

```
afam <- (w$活跃保险箱余额+w$活跃身上携带)/w$月活跃用户数
```

```
slac <- round(smpc/afam,2)
```

```
unusual <- ifelse(w$日活跃.月活跃>0.2,"正常","异常")
```

```
u <- data.frame(cbind(cadr,slac,unusual))
```

```
u$cadr <- as.numeric(u$cadr)
```

```
u$slac <- as.numeric(u$slac)
```

# 建立决策树模型

```
library(rpart)
```

```
r<-rpart(unusual ~ .,data=u,method="class",parms=list(split="information"))
```

```
rattle::fancyRpartPlot(r)
```

