

# Game\_analysis\_313

## 问题一：留存问题。

## 用户留存与渠道关系分析

游戏为了获取新用户，一般都会在外部渠道进行资源投放，例如微信，QQ，搜索引擎，游戏门户网站等等。外部渠道的投放都是需要真金白银的支持，所以对获客渠道的分析至关重要，它直接决定我们能否将一定的预算效果最大化。所以我们需要做一个不同渠道引流的玩家跟留存率的关系分析，以确定企业的资源得到了最大化的利用。

## dataset overview

```
setwd("~/Documents/GitHub/Game_DataMining_With_R/nino/project1")  
library(readr)  
library(caret)
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
login <- read.csv("/Users/nino/Documents/GitHub/Game_DataMining_With_R/nino/project1/  
logindata.csv", fileEncoding = "GBK")  
str(login)  
summary(login)  
head(login)  
dim(login)  
colnames(login)
```

从数据集的总结来看，该数据集包括30000条数据，7个变量。分别是：“渠道名称”“游戏名称”“是否付费”“性别”“年龄”“登录天数”“登录次数”。

另外需要注意的是，除了年龄变量以外其他变量应为factor类变量，所以进行相应的转换。

```
dummy <- dummyVars(~ ., data = login)  
dummyTsrf <- data.frame(predict(dummy, newdata = login))
```

通过函数求出变量间的相关系数。

```

cor.prob <- function (X, dfr = nrow(X) - 2) {
  R <- cor(X, use="pairwise.complete.obs")
  above <- row(R) < col(R)
  r2 <- R[above]^2
  Fstat <- r2 * dfr/(1 - r2)
  R[above] <- 1 - pf(Fstat, 1, dfr)
  R[row(R) == col(R)] <- NA
  R
}
flattenSquareMatrix <- function(m) {
  if( (class(m) != "matrix") | (nrow(m) != ncol(m))) stop("Must be a square matrix.")
  if(!identical(rownames(m), colnames(m))) stop("Row and column names must be equal.")
)
  ut <- upper.tri(m)
  data.frame(i = rownames(m)[row(m)[ut]],
            j = rownames(m)[col(m)[ut]],
            cor=t(m)[ut],
            p=m[ut])
}
corMasterList <- flattenSquareMatrix(cor.prob(dummyTsrfr))
corList <- corMasterList[order(-abs(corMasterList$cor)), ]
print(head(corList, 20))

```

##	i	j	cor	p
## 276	是否付费.否	是否付费.是	-1.0000000	0
## 325	性别.女	性别.男	-1.0000000	0
## 562	渠道名称.渠道A	登录次数.1次	0.9984244	0
## 495	登录天数.3天	登录天数.5天以上	-0.7887687	0
## 189	游戏名称.游戏C	游戏名称.游戏E	-0.6439172	0
## 345	游戏名称.游戏E	年龄	-0.5335248	0
## 249	游戏名称.游戏C	是否付费.否	-0.4442365	0
## 271	游戏名称.游戏C	是否付费.是	0.4442365	0
## 779	登录次数.31至40次	登录次数.51至60次	-0.4431728	0
## 494	登录天数.2天	登录天数.5天以上	-0.4390142	0
## 294	游戏名称.游戏C	性别.女	-0.4320280	0
## 318	游戏名称.游戏C	性别.男	0.4320280	0
## 700	登录次数.11至20次	登录次数.31至40次	-0.3971948	0
## 667	渠道名称.渠道A	登录次数.31至40次	-0.3693758	0
## 701	登录次数.1次	登录次数.31至40次	-0.3687938	0
## 152	游戏名称.游戏A	游戏名称.游戏C	-0.3665339	0
## 251	游戏名称.游戏E	是否付费.否	0.3162397	0
## 273	游戏名称.游戏E	是否付费.是	-0.3162397	0
## 343	游戏名称.游戏C	年龄	0.3139263	0
## 818	登录次数.31至40次	登录次数.61.100次	-0.3078920	0

通过计算发现,

1, 渠道A与登陆次数一次的相关系数很大, 初步怀疑渠道A的用户黏性较差。但因为相关系数并不一定代表着因果关系, 还需进一步分析。

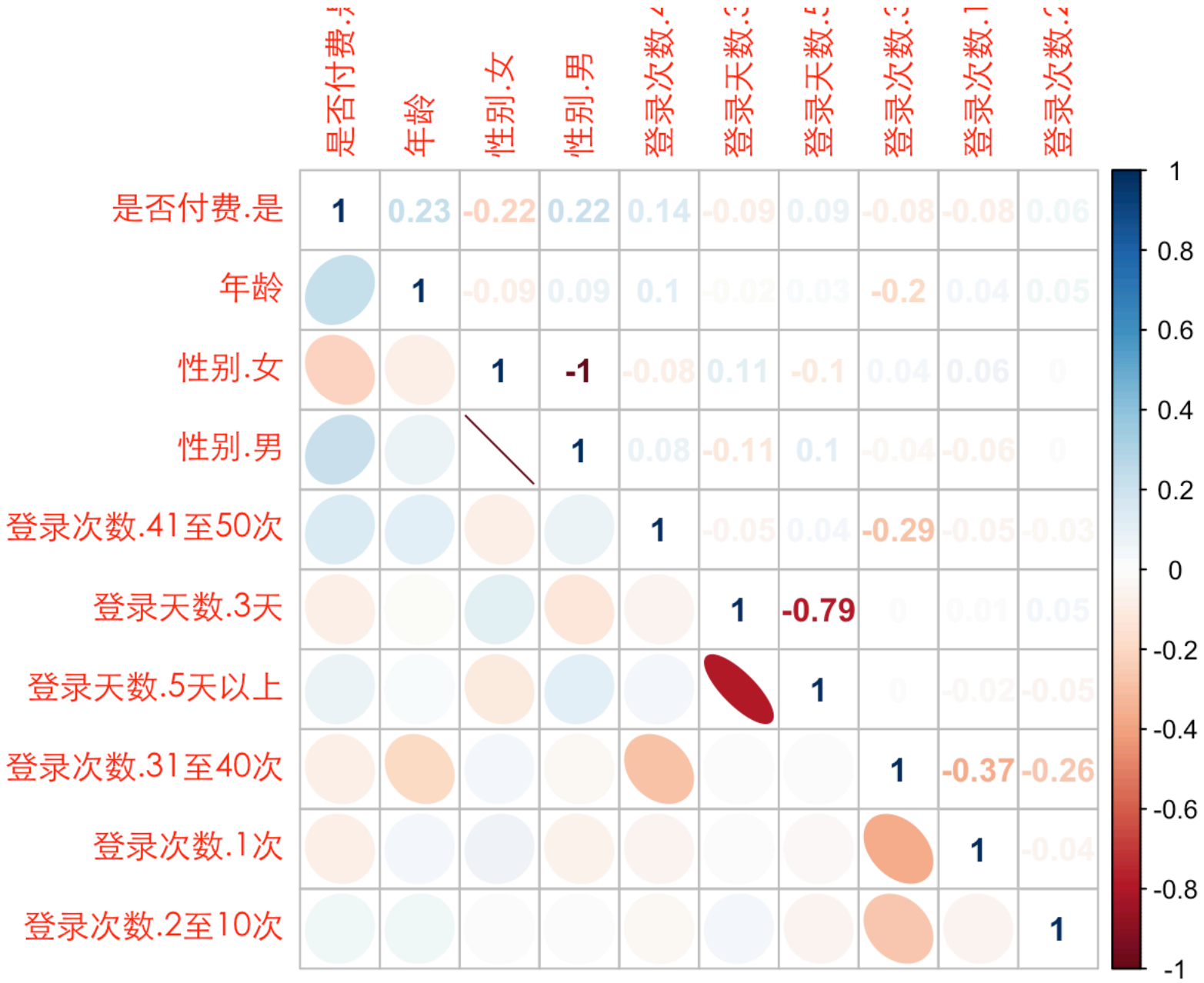
- 2, 游戏E与年龄有负相关关系, 说明游戏E的游戏用户年龄较小。
- 3, 游戏C与用户是否付费有正相关关系, 说明游戏C的玩家付费意愿相对较高。

# 可视化

```
paycor <- subset(corList, (abs(cor) > 0.04 & i %in% c("是否付费.是")))  
sub <- as.character(paycor$j)  
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
plotcor <- corrplot.mixed(cor(dummyTsrf[, c("是否付费.是", sub)]),  
  lower = "ellipse", upper = "number", tl.pos = "lt", diag = "u", family='STXihei')
```



通过绘制与“是否付费.是”相关系数大于0.4的图发现

- 1, 男性比女性更容易付费。
- 2, 年龄越大的玩家越容易付费。

3，登录次数越多越容易付费。

## 问题二： 个性化推荐。

# 不同玩家的消费偏好分析。

之前在面试中提到了对玩家的行为数据进行分析并做个性化推荐。当时我回答了如何做第一步，用户的分类，这里我想展现一下为了支持产品研发，后续能做的一些分析研究。

在对玩家进行分类以后我们可以对不同类型的玩家的消费趋向做分析以达到精细化推荐的目的。例如给喜欢买金币的玩家推送金币活动，喜欢刷喇叭的玩家推送喇叭礼包等。

## dataset overview

```
library(ca)
playershop <- read.csv("/Users/nino/Documents/GitHub/Game_DataMining_With_R/nino/project1/玩家喜好分析数据.csv", T)
rownames(playershop) <- playershop[,1]
summary(playershop)
```

##	道具名称	A类玩家	B类玩家	C类玩家
##	20000个金币	: 1 Min. : 0.0	Min. : 0.00	Min. : 125.0
##	20个钻	: 1 1st Qu.: 3.0	1st Qu.: 46.25	1st Qu.: 240.5
##	2个钻	: 1 Median : 5.5	Median : 117.50	Median : 313.5
##	50000个金币	: 1 Mean : 25.3	Mean : 395.30	Mean :1257.0
##	50个钻	: 1 3rd Qu.: 9.0	3rd Qu.: 168.25	3rd Qu.: 391.0
##	5个大师精灵球	: 1 Max. :400.0	Max. :4600.00	Max. :7500.0
##	(Other)	:14		
##	D类玩家	E类玩家		
##	Min. : 31.0	Min. : 142.0		
##	1st Qu.: 92.0	1st Qu.: 195.0		
##	Median :149.5	Median : 253.5		
##	Mean :213.4	Mean : 385.5		
##	3rd Qu.:200.0	3rd Qu.: 293.5		
##	Max. :960.0	Max. :2270.0		
##				

```
head(playershop)
```

	道具名称 <fctr>	A类玩家 <int>	B类玩家 <int>	C类玩家 <int>	D类玩家 <int>	E类玩家 <int>
8个钻	8个钻	4	181	391	36	2270
感恩大礼包	感恩大礼包	2	124	314	200	285

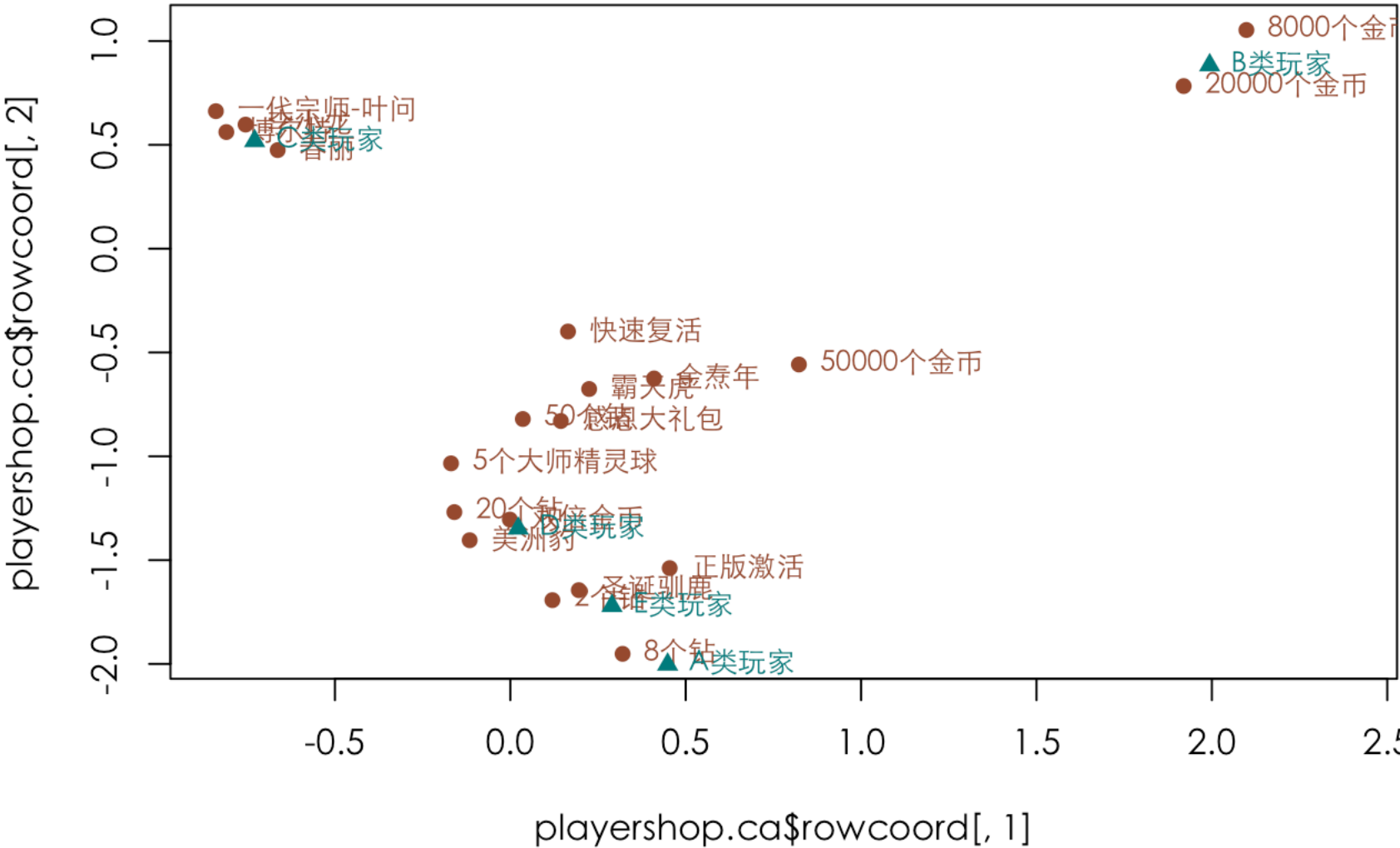
快速复活	快速复活	10	136	313	125	166
5个大师精灵球	5个大师精灵球	3	15	257	89	247
圣诞驯鹿	圣诞驯鹿	3	101	125	800	229
双倍金币	双倍金币	3	38	218	247	260
6 rows						

通过数据集总计，我们可以很直观的看到该数据集是一个不同类型玩家（5类）在游戏对不同道具(20类）的消费数据。

## 可视化

```
playershop.ca <- ca(playershop[, -1])
plot(playershop.ca$rowcoord[,1],playershop.ca$rowcoord[,2],pch = 16,col = rgb(0.6,0.3,0.2),
      xlim = c(min(playershop.ca$rowcoord[,1]),max(playershop.ca$rowcoord[,1])+0.3),
      main = "玩家购买物品偏好分析" ,family = "STXihei")
text(playershop.ca$rowcoord[,1],playershop.ca$rowcoord[,2],labels = playershop.ca$row
names,
      cex = 0.8,pos = 4,col = rgb(0.6,0.3,0.2), family = "STXihei")
points(playershop.ca$colcoord[,1],playershop.ca$colcoord[,2],pch = 17,col = "#007e7e"
, family = "STXihei")
text(playershop.ca$colcoord[,1],playershop.ca$colcoord[,2],labels = playershop.ca$col
names,
      cex = 0.8,pos = 4,col = "#007e7e",family = "STXihei")
```

## 玩家购买物品偏好分析



通过简单的对应分析，绘图，我们可以很直观的看到不同类型的玩家喜欢购买的物品偏好。如C类玩家喜欢买角色，B类玩家喜欢买金币等，为之后的个性化推荐打下基础。