CRM Report

Hyeonho Lee

2018년 8월 25일

## Load Data

# library(dplyr)  
# 데이터 조작에 필요한 패키지  
mailorder = read.csv('D:/second semester/crm/mailorder.csv')  
# load data

## Question 1.

mailorder1 = head(mailorder, 2000)  
mailorder2 = tail(mailorder, 2000)  
# 2000개의 estimation sample과 validation sample로 나눔  
  
ind = sample(1:nrow(mailorder2), nrow(mailorder2), replace = FALSE)  
mailorder2\_1 = mailorder2[ind <= 500,]  
#   
paste0('500명 중 실제로 구매한 인원은 ', sum(mailorder2\_1$purchase == 1), '명 입니다.')

## [1] "500명 중 실제로 구매한 인원은 27명 입니다."

## Question 2.

mailorder1$R = cut(mailorder1$recency, breaks = c(min(mailorder1$recency), 12,   
 max(mailorder1$recency)),include.lowest = T, labels = c(2,1))  
mailorder1$F = cut(mailorder1$frequency, breaks = c(min(mailorder1$frequency)-1, 2,   
 max(mailorder1$frequency)),rigth = T, labels = c(1, 2))  
mailorder1$M = cut(mailorder1$monetary, breaks = c(min(mailorder1$monetary)-1, 208,   
 max(mailorder1$monetary)),rigth = T, labels = c(1, 2))  
mailorder2$R = cut(mailorder2$recency, breaks = c(min(mailorder2$recency), 12,   
 max(mailorder2$recency)),include.lowest = T, labels = c(2,1))  
mailorder2$F = cut(mailorder2$frequency, breaks = c(min(mailorder2$frequency)-1, 2,   
 max(mailorder2$frequency)),rigth = T, labels = c(1, 2))  
mailorder2$M = cut(mailorder2$monetary, breaks = c(min(mailorder2$monetary)-1, 208,   
 max(mailorder2$monetary)),rigth = T, labels = c(1, 2))  
# 2X2X2 RFM codes 에 따라 estimation and validation sample을 분류함  
  
mailorder1$R = varhandle::unfactor(mailorder1$R)  
mailorder1$F = varhandle::unfactor(mailorder1$F)  
mailorder1$M = varhandle::unfactor(mailorder1$M)  
mailorder2$R = varhandle::unfactor(mailorder2$R)  
mailorder2$F = varhandle::unfactor(mailorder2$F)  
mailorder2$M = varhandle::unfactor(mailorder2$M)  
# 분류하는 과정에서 factor형 변수로 변환 되었기 때문에 factor를 풀어주는 코드임  
  
mean\_pur = mailorder1 %>% select(R, F, M, purchase) %>% group\_by(R, F, M) %>%   
 summarise(mean\_purchase = mean(purchase))%>% arrange(desc(mean\_purchase))  
knitr::kable(mean\_pur, caption = 'A caption')

A caption

|  |  |  |  |
| --- | --- | --- | --- |
| R | F | M | mean\_purchase |
| 2 | 2 | 1 | 0.1751825 |
| 2 | 2 | 2 | 0.1672131 |
| 2 | 1 | 2 | 0.0891473 |
| 2 | 1 | 1 | 0.0636132 |
| 1 | 2 | 2 | 0.0611354 |
| 1 | 2 | 1 | 0.0504202 |
| 1 | 1 | 2 | 0.0465116 |
| 1 | 1 | 1 | 0.0290698 |

# mailorder1 즉 R,F,M을 기준으로 estimation sample의 평균 mean\_purchase 계산  
  
for( i in 1:nrow(mailorder2))  
{  
 for( j in 1:nrow(mean\_pur))  
 {  
 if(mean\_pur$R[j] == mailorder2$R[i] &  
 mean\_pur$F[j] == mailorder2$F[i] &  
 mean\_pur$M[j] == mailorder2$M[i])  
 {  
 mailorder2$mean\_purchase[i] = mean\_pur$mean\_purchase[j]  
 }  
 }   
}  
# mailorder2(validation sample에 R,F,M 부여)의 R,F,M이 mailorder1의 R,F,M과 같을 때,  
# mailorder2에 mean\_purchase를 추가함  
  
mailorder2 %>% select(purchase, mean\_purchase) %>%   
 arrange(desc(mean\_purchase)) %>% head(500) %>% summarise(mean = mean(purchase))

## mean  
## 1 0.152

# mean\_purchase가 추가된 mailorder2를 purchase와 mean\_purchase만 선택하여,  
# mean\_purchase를 기준으로 내림차순하여, 상위 500개의 purchase확률을 보여줌  
  
paste0('500명 중 실제로 구매한 인원은 ', mailorder2 %>% select(purchase, mean\_purchase) %>%   
 arrange(desc(mean\_purchase)) %>% head(500) %>% filter(purchase == 1) %>% nrow(),  
 '명 입니다.')

## [1] "500명 중 실제로 구매한 인원은 76명 입니다."

## Question 3.

mailorder1$R = cut(mailorder1$recency, breaks = c(min(mailorder1$recency),   
 4, 8, 12, 16, max(mailorder1$recency)),include.lowest = T,   
 labels = c(5,4,3,2,1))  
mailorder1$F = cut(mailorder1$frequency, breaks = c(min(mailorder1$frequency)-1,   
 1, 2, 5, 9, max(mailorder1$frequency)),rigth = T,   
 labels = c(1,2,3,4,5))  
mailorder1$M = cut(mailorder1$monetary, breaks = c(min(mailorder1$monetary)-1,   
 113, 181, 242, 299, max(mailorder1$monetary)),rigth = T,   
 labels = c(1,2,3,4,5))  
mailorder2$R = cut(mailorder2$recency, breaks = c(min(mailorder2$recency),   
 4, 8, 12, 16, max(mailorder2$recency)),include.lowest = T,   
 labels = c(5,4,3,2,1))  
mailorder2$F = cut(mailorder2$frequency, breaks = c(min(mailorder2$frequency)-1,   
 1, 2, 5, 9, max(mailorder2$frequency)),rigth = T,   
 labels = c(1,2,3,4,5))  
mailorder2$M = cut(mailorder2$monetary, breaks = c(min(mailorder2$monetary)-1,   
 113, 181, 242, 299, max(mailorder2$monetary)),rigth = T,   
 labels = c(1,2,3,4,5))  
# 5X5X5 RFM codes 에 따라 estimation and validation sample을 분류함  
  
mailorder1$R = varhandle::unfactor(mailorder1$R)  
mailorder1$F = varhandle::unfactor(mailorder1$F)  
mailorder1$M = varhandle::unfactor(mailorder1$M)  
mailorder2$R = varhandle::unfactor(mailorder2$R)  
mailorder2$F = varhandle::unfactor(mailorder2$F)  
mailorder2$M = varhandle::unfactor(mailorder2$M)  
# 분류하는 과정에서 factor형 변수로 변환 되었기 때문에 factor를 풀어주는 코드임  
  
mean\_pur = mailorder1 %>% select(R, F, M, purchase) %>% group\_by(R, F, M) %>%   
 summarise(mean\_purchase = mean(purchase))%>% arrange(desc(mean\_purchase))  
knitr::kable(mean\_pur, caption = 'A caption')

A caption

|  |  |  |  |
| --- | --- | --- | --- |
| R | F | M | mean\_purchase |
| 4 | 5 | 4 | 0.6000000 |
| 4 | 1 | 5 | 0.5000000 |
| 4 | 4 | 2 | 0.5000000 |
| 4 | 5 | 2 | 0.5000000 |
| 5 | 5 | 2 | 0.4000000 |
| 5 | 3 | 2 | 0.3750000 |
| 4 | 3 | 4 | 0.3333333 |
| 2 | 3 | 4 | 0.3000000 |
| 3 | 4 | 3 | 0.3000000 |
| 5 | 4 | 5 | 0.2962963 |
| 4 | 5 | 3 | 0.2857143 |
| 5 | 3 | 5 | 0.2857143 |
| 3 | 4 | 2 | 0.2500000 |
| 5 | 2 | 3 | 0.2500000 |
| 5 | 5 | 3 | 0.2500000 |
| 5 | 5 | 5 | 0.2500000 |
| 3 | 5 | 5 | 0.2272727 |
| 4 | 5 | 5 | 0.2272727 |
| 3 | 5 | 4 | 0.2000000 |
| 4 | 1 | 1 | 0.2000000 |
| 4 | 4 | 3 | 0.2000000 |
| 3 | 2 | 4 | 0.1785714 |
| 2 | 3 | 2 | 0.1764706 |
| 4 | 2 | 2 | 0.1578947 |
| 2 | 4 | 4 | 0.1538462 |
| 3 | 3 | 3 | 0.1538462 |
| 5 | 4 | 3 | 0.1538462 |
| 4 | 4 | 5 | 0.1428571 |
| 5 | 1 | 5 | 0.1428571 |
| 5 | 3 | 3 | 0.1428571 |
| 2 | 3 | 1 | 0.1333333 |
| 1 | 1 | 4 | 0.1304348 |
| 2 | 2 | 3 | 0.1304348 |
| 5 | 2 | 1 | 0.1304348 |
| 1 | 3 | 3 | 0.1250000 |
| 4 | 3 | 3 | 0.1250000 |
| 5 | 2 | 4 | 0.1250000 |
| 3 | 4 | 4 | 0.1176471 |
| 5 | 1 | 2 | 0.1176471 |
| 2 | 4 | 5 | 0.1111111 |
| 3 | 3 | 4 | 0.1111111 |
| 3 | 5 | 3 | 0.1111111 |
| 4 | 2 | 5 | 0.1111111 |
| 2 | 1 | 4 | 0.1034483 |
| 5 | 1 | 1 | 0.1034483 |
| 2 | 3 | 3 | 0.1000000 |
| 4 | 1 | 2 | 0.0952381 |
| 4 | 4 | 4 | 0.0909091 |
| 2 | 2 | 2 | 0.0833333 |
| 3 | 1 | 4 | 0.0833333 |
| 3 | 1 | 2 | 0.0731707 |
| 1 | 5 | 3 | 0.0714286 |
| 1 | 5 | 4 | 0.0714286 |
| 3 | 2 | 2 | 0.0645161 |
| 1 | 4 | 5 | 0.0625000 |
| 4 | 3 | 2 | 0.0625000 |
| 5 | 1 | 4 | 0.0625000 |
| 5 | 2 | 5 | 0.0625000 |
| 4 | 1 | 4 | 0.0476190 |
| 4 | 2 | 4 | 0.0454545 |
| 3 | 4 | 5 | 0.0434783 |
| 1 | 2 | 1 | 0.0384615 |
| 1 | 5 | 5 | 0.0370370 |
| 2 | 2 | 4 | 0.0370370 |
| 1 | 1 | 3 | 0.0357143 |
| 1 | 2 | 2 | 0.0344828 |
| 3 | 2 | 3 | 0.0344828 |
| 1 | 2 | 3 | 0.0333333 |
| 2 | 5 | 5 | 0.0312500 |
| 2 | 2 | 1 | 0.0263158 |
| 2 | 1 | 2 | 0.0256410 |
| 3 | 1 | 1 | 0.0250000 |
| 2 | 1 | 1 | 0.0212766 |
| 1 | 1 | 1 | 0.0000000 |
| 1 | 1 | 2 | 0.0000000 |
| 1 | 1 | 5 | 0.0000000 |
| 1 | 2 | 4 | 0.0000000 |
| 1 | 2 | 5 | 0.0000000 |
| 1 | 3 | 1 | 0.0000000 |
| 1 | 3 | 2 | 0.0000000 |
| 1 | 3 | 4 | 0.0000000 |
| 1 | 3 | 5 | 0.0000000 |
| 1 | 4 | 1 | 0.0000000 |
| 1 | 4 | 2 | 0.0000000 |
| 1 | 4 | 3 | 0.0000000 |
| 1 | 4 | 4 | 0.0000000 |
| 1 | 5 | 2 | 0.0000000 |
| 2 | 1 | 3 | 0.0000000 |
| 2 | 1 | 5 | 0.0000000 |
| 2 | 2 | 5 | 0.0000000 |
| 2 | 3 | 5 | 0.0000000 |
| 2 | 4 | 1 | 0.0000000 |
| 2 | 4 | 2 | 0.0000000 |
| 2 | 4 | 3 | 0.0000000 |
| 2 | 5 | 2 | 0.0000000 |
| 2 | 5 | 3 | 0.0000000 |
| 2 | 5 | 4 | 0.0000000 |
| 3 | 1 | 3 | 0.0000000 |
| 3 | 1 | 5 | 0.0000000 |
| 3 | 2 | 1 | 0.0000000 |
| 3 | 2 | 5 | 0.0000000 |
| 3 | 3 | 1 | 0.0000000 |
| 3 | 3 | 2 | 0.0000000 |
| 3 | 3 | 5 | 0.0000000 |
| 3 | 4 | 1 | 0.0000000 |
| 3 | 5 | 2 | 0.0000000 |
| 4 | 1 | 3 | 0.0000000 |
| 4 | 2 | 1 | 0.0000000 |
| 4 | 2 | 3 | 0.0000000 |
| 4 | 3 | 1 | 0.0000000 |
| 4 | 3 | 5 | 0.0000000 |
| 4 | 4 | 1 | 0.0000000 |
| 5 | 1 | 3 | 0.0000000 |
| 5 | 2 | 2 | 0.0000000 |
| 5 | 3 | 1 | 0.0000000 |
| 5 | 3 | 4 | 0.0000000 |
| 5 | 4 | 1 | 0.0000000 |
| 5 | 4 | 2 | 0.0000000 |
| 5 | 4 | 4 | 0.0000000 |
| 5 | 5 | 4 | 0.0000000 |

# mailorder1 즉 R,F,M을 기준으로 estimation sample의 평균 mean\_purchase 계산  
  
  
for( i in 1:nrow(mailorder2))  
{  
 for( j in 1:nrow(mean\_pur))  
 {  
 if(mean\_pur$R[j] == mailorder2$R[i] &  
 mean\_pur$F[j] == mailorder2$F[i] &  
 mean\_pur$M[j] == mailorder2$M[i])  
 {  
 mailorder2$mean\_purchase[i] = mean\_pur$mean\_purchase[j]  
 }  
 }   
}  
# mailorder2(validation sample에 R,F,M 부여)의 R,F,M이 mailorder1의 R,F,M과 같을 때,  
# mailorder2에 mean\_purchase를 추가함  
  
mailorder2 %>% select(purchase, mean\_purchase) %>%   
 arrange(desc(mean\_purchase)) %>% head(500) %>% summarise(mean = mean(purchase))

## mean  
## 1 0.124

# mean\_purchase가 추가된 mailorder2를 purchase와 mean\_purchase만 선택하여,  
# mean\_purchase를 기준으로 내림차순하여, 상위 500개의 purchase확률을 보여줌  
  
paste0('500명 중 실제로 구매한 인원은 ', mailorder2 %>% select(purchase, mean\_purchase) %>%   
 arrange(desc(mean\_purchase)) %>% head(500) %>% filter(purchase == 1) %>% nrow(),  
 '명 입니다.')

## [1] "500명 중 실제로 구매한 인원은 62명 입니다."

## Question 4.

model = lm(purchase~recency+frequency+monetary, mailorder1)  
pred = predict(model, mailorder2)  
  
mailorder2$predict = pred  
  
mailorder2 %>% select(purchase, recency, frequency, monetary, predict) %>%   
 arrange(desc(predict)) %>% head(500) %>% summarise(mean = mean(purchase))

## mean  
## 1 0.16

## Question 5.