RFM

GS 홈쇼핑

Table of Contents

# 1. 고객별 등급 매기기 (dplyr)

library(readxl)  
library(dplyr)  
dataset <- read\_excel("retail.xlsx")  
# head(dataset, 2)  
cus\_dataset <- dataset %>%   
 group\_by(`Customer Name`) %>%  
 summarise(Recent = max(`Order Date`),  
 Freq = n(),  
 Money = sum(Sales))  
cus\_dataset <- cus\_dataset %>%   
 mutate(R\_score = ntile(Recent, 5),  
 F\_score = ntile(Freq, 5),  
 M\_score = ntile(Money, 5)) %>%   
 arrange(desc(Recent), desc(Freq), desc(Money)) %>%   
 select(`Customer Name`,   
 R\_score, F\_score, M\_score,   
 Recent, Freq, Money)   
head(cus\_dataset, 2)

## # A tibble: 2 x 7  
## `Customer Name` R\_score F\_score M\_score Recent Freq Money  
## <chr> <int> <int> <int> <dttm> <int> <dbl>  
## 1 Chuck Clark 5 5 4 2014-12-31 00:00:00 19 2869  
## 2 Patrick O'Donnell 5 3 3 2014-12-31 00:00:00 13 2494

총 9994개의 판매 기록을 집계하여 793의 고객에 대해서 스코어링을 하였습니다.

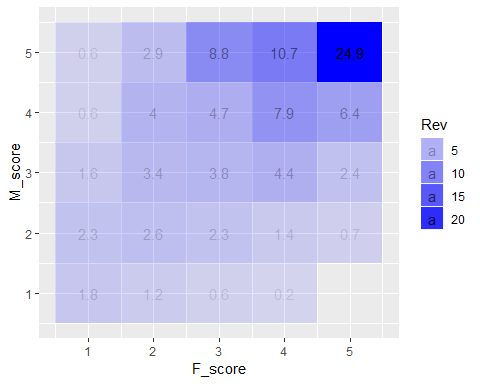
# 2. 파레토 분석 - 20% 고객이 80%의 매출을 차지하는가? (dplyr)

pareto\_M\_score <- cus\_dataset %>%   
 group\_by(M\_score) %>%  
 summarise(tot\_money = sum(Money)) %>%   
 arrange(desc(tot\_money))  
pareto\_M\_score <- pareto\_M\_score %>%  
 mutate(proportion = tot\_money/sum(tot\_money))   
pareto\_M\_score

## # A tibble: 5 x 3  
## M\_score tot\_money proportion  
## <int> <dbl> <dbl>  
## 1 5 1101851 0.480   
## 2 4 539957 0.235   
## 3 3 357936 0.156   
## 4 2 213073 0.0927  
## 5 1 84537 0.0368

# 3. 간단한 차트 그리기 (ggplot2)

tot\_rev <- sum(cus\_dataset$Money)  
rev\_anal <- cus\_dataset %>% group\_by(F\_score, M\_score) %>%  
 summarise(Rev = round(100\*sum(Money)/tot\_rev, 1))  
library(ggplot2)  
ggplot(rev\_anal, aes(x = F\_score, y = M\_score, alpha = Rev)) +   
 geom\_tile(fill = "blue") + geom\_text(aes(label = Rev))



# 4. RFM 분석이란 무엇인가?

RFM 분석은 고객별로 3개의 지표(R,M,F)를 기준으로 5분위로 나누어 점수를 매겨 고객을 분류하고 마켓팅 전략에 활용하는 마켓팅 분석 기법이다.

* R: Recency (최근 구매일)
* F: Frequency (총 구매횟수)
* M: Monetary (총 구매액)

일반적으로 사용되는 고객 분류와 는 아래와 같다.

|  |  |  |  |
| --- | --- | --- | --- |
| Segment | R\_score | mean(F\_score, M\_score) |  |
| Champions | 4-5 | 4-5 |  |
| Loyal | 2-5 | 3-5 |  |
| Potential Loyalist | 3-5 | 1-3 | Offer loyalty program, recommend products |
| **Recent Customers** | 4-5 | 1-2 | Provide support, build relationship |
| Promising | 3-4 | 1-2 | Create brand awareness, offer free trials |
| Customers Needing Attention | 2-3 | 2-3 | Recommend based on previous purchase, reactivate |
| About To Sleep | 2-3 | 1-2 | Recomment popular products, reconnect with them |
| At Risk | 1-2 | 2-5 | Personalized email, offer renewals, provide resources |
| Can’t Lose Them | 1-2 | 4-5 | Don’t lose them to competition, talk to them |
| Hibernating | 1-2 | 1-2 | Offer other relevant products and special discounts |
| Lost | 1 | 1 | Reach out compaign, ignore otherwise |

# 5. Recent Customers만 조회하기

cus\_dataset %>%   
 mutate(FM\_avg = pmin(F\_score, M\_score)) %>%  
 filter(R\_score >= 4) %>%  
 filter(FM\_avg <= 2)

## # A tibble: 133 x 8  
## `Customer Name` R\_score F\_score M\_score Recent Freq Money  
## <chr> <int> <int> <int> <dttm> <int> <dbl>  
## 1 Jill Matthias 5 1 1 2014-12-31 00:00:00 7 305  
## 2 Erica Bern 5 1 2 2014-12-31 00:00:00 6 1644  
## 3 Ken Brennan 5 2 2 2014-12-30 00:00:00 9 984  
## 4 Erica Hernandez 5 3 2 2014-12-29 00:00:00 11 1220  
## 5 Lori Olson 5 2 1 2014-12-29 00:00:00 9 644  
## 6 Pamela Coakley 5 1 3 2014-12-29 00:00:00 5 1832  
## 7 Greg Hansen 5 1 1 2014-12-29 00:00:00 3 147  
## 8 Karen Bern 5 2 4 2014-12-27 00:00:00 10 3152  
## 9 Tracy Collins 5 3 1 2014-12-27 00:00:00 10 742  
## 10 Yoseph Carroll 5 2 5 2014-12-27 00:00:00 8 5455  
## # ... with 123 more rows, and 1 more variable: FM\_avg <int>

Reference: <https://www.putler.com/rfm-analysis/>