Bookbinders Book Club

Jordi López Sintas

3/19/2018

Los datos

The case Bookbinder Book Club, BBBClub, illustrates the use of three ways that can be use to select the potential clients that will be the target of our marketing efforts: (1) the traditional RFM model, (2) the linear model, and (3) the logistic model.

We need two data sets, one historic data set with the past behavior of a sample of clients, and another data set with the data of possible new clients of a book, in this case.

The historic data set with the past behavior of a sample of clients

```
BBBClub.choice<-read.delim(
  "R-Bookbinders Book Club Data (Customer Choice).txt")
BBBClub.holdout<-read.delim(
  "R-Bookbinders Book Club Data (Customer Choice) Holdout Sample.txt")

knitr::kable(
  t(head(BBBClub.choice)),
  digits=2,
  caption = "Descripción de las 6 primeras empresas de los datos históricos")
```

Table 1: Descripción de las 6 primeras empresas de los datos históricos

	1	2	3	4	5	6
choice	1	1	1	1	1	1
gender	1	1	1	1	0	1
amount	113	418	336	180	320	268
freq	8	6	18	16	2	4
last	1	11	6	5	3	1
first	8	66	32	42	18	4
child	0	0	2	2	0	0
youth	1	2	0	0	0	0
cook	0	3	1	0	0	0
diy	0	2	1	1	1	0
art	0	3	2	1	2	0

The data set with prospects

```
knitr::kable(
  t(head(BBBClub.holdout)),
  digits=2,
  caption = "Descripción de las 6 primeras empresas de los nuevos datos ")
```

Table 2: Descripción de las 6 primeras empresas de los nuevos datos

	1	2	3	4	5	6
choice	1	1	1	1	1	1
gender	0	1	1	0	1	0
amount	287	215	261	24	120	66
freq	12	4	2	4	8	2
last	4	1	1	1	1	4
first	24	4	2	4	8	16
child	0	0	0	1	0	0
youth	3	0	0	0	0	0
cook	0	0	0	0	0	1
diy	0	0	0	0	0	1
art	1	1	1	0	1	1

RFM model

The RFM is based on the Recency, Frequency and Monetary value of old clients with similar characteristics.

- Recency: Tiempo transcurrido desde la última compra ($<$ 2 wks, 2-6 wks, 7-8 wks, and $>$ 12 wks)
- Frequency: Número de compras realizadas desde una determinada fecha hasta la actualidad (10, 8, 6, 3, 1)
- Monetary Value: Valor de las compras desde una determina fecha ($>$ 500, 300-500, 200-300, 100-200, $<$ 200)

First we apply the RFM model to the new data (prospects)

Asignar una puntuación a R, F y M según la experiencia. Por ejemplo:

Recency:

últimos 3meses, 25 puntos 3-6 meses, 20 puntos 6-9 meses, 10 putnos 10-18 meses, 5, puntos Más de 18 meses, 0 puntos

Aplicar una regla similar para *Frequency* y *Monetary*

Finalmente formamos una puntuación en RFM (*scores*) sumándolos (se podrían ponderar según la importancia de *Recency*, *Frequency* y *Monetary Value*)

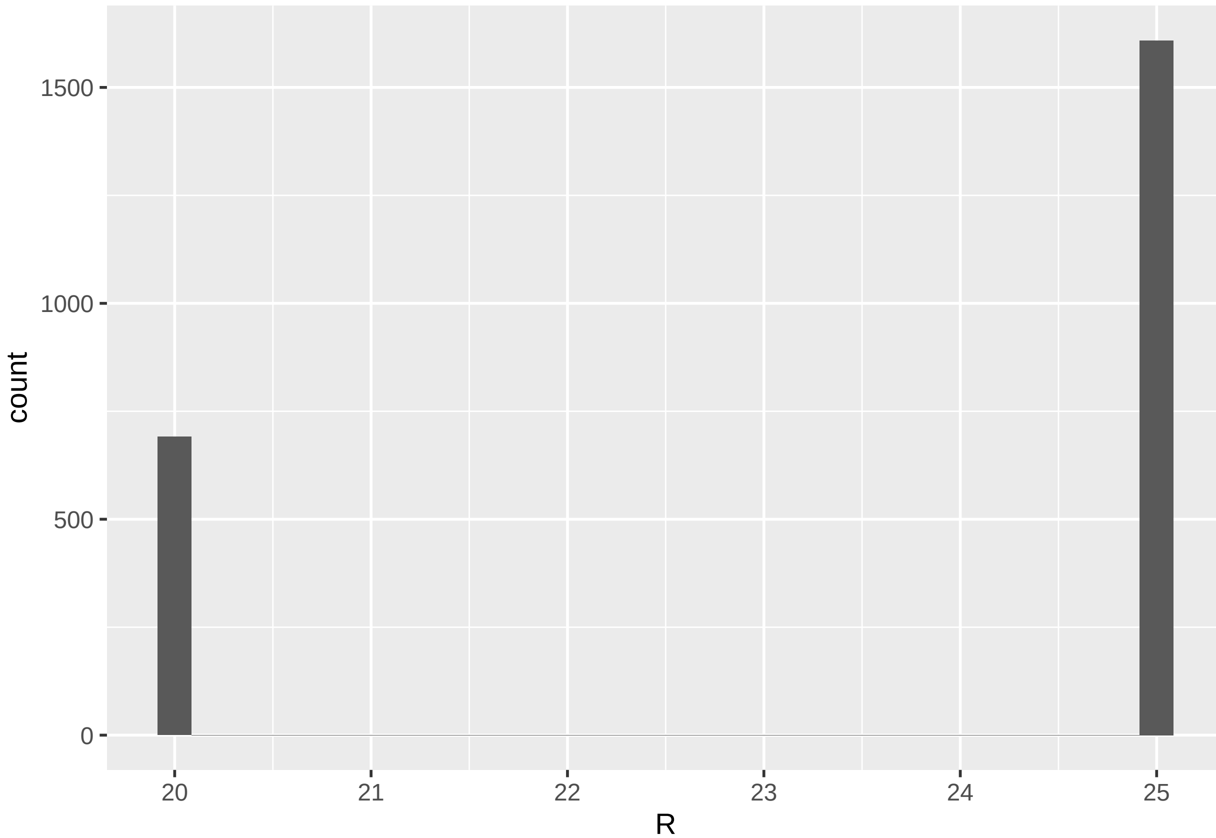
Esto es lo que hacemos con la función RFM que tenemos en el guión `marketing-models.R`.

Table 3: Descripción de las 6 primeras empresas de los nuevos datos

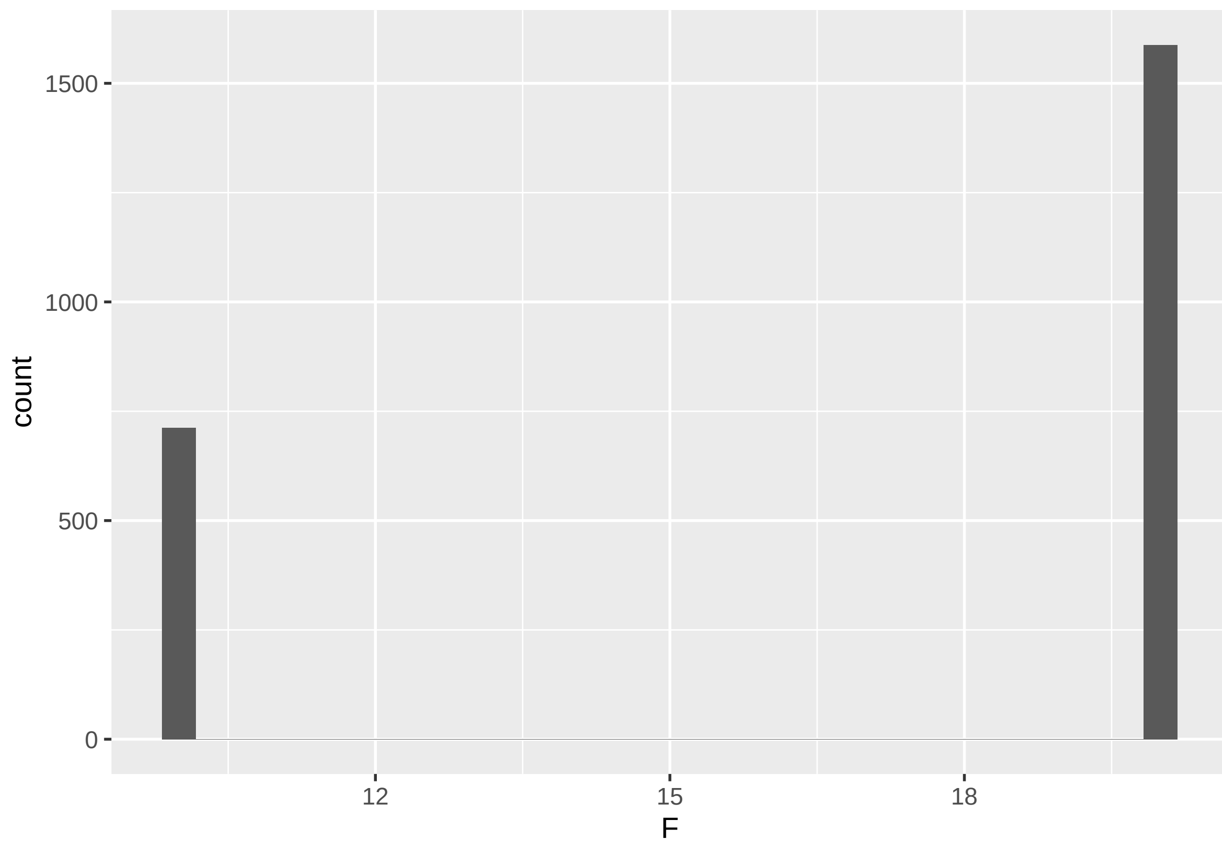
	1	2	3	4	5	6
choice	1	1	1	1	1	1
gender	0	1	1	0	1	0
amount	287	215	261	24	120	66
freq	12	4	2	4	8	2
last	4	1	1	1	1	4
first	24	4	2	4	8	16
child	0	0	0	1	0	0
youth	3	0	0	0	0	0
cook	0	0	0	0	0	1
diy	0	0	0	0	0	1
art	1	1	1	0	1	1
R	20	25	25	25	25	20
F	20	10	10	10	10	10
M	40	30	40	10	20	20
RFM	80	65	75	45	55	50

Plots

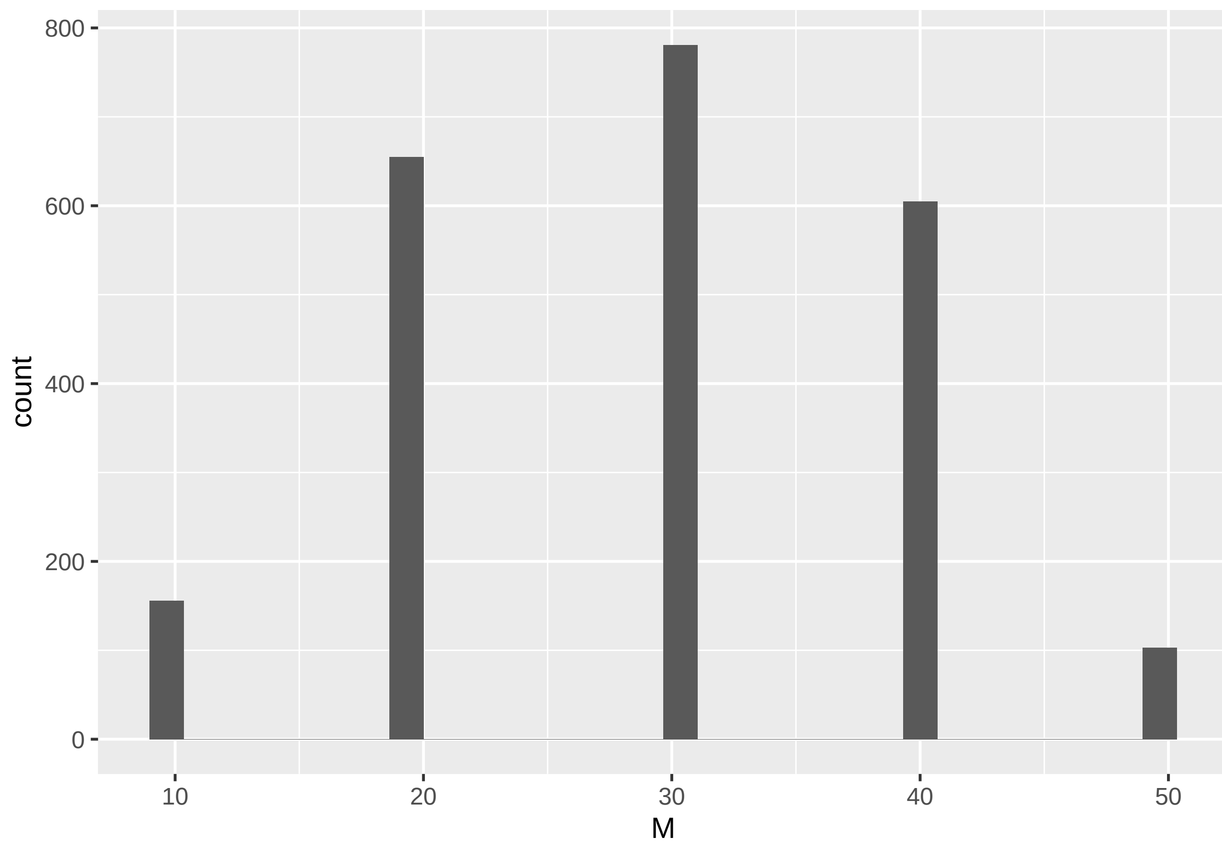
```
library(ggplot2)
ggplot(BBBClub.holdout, aes(R)) +
  geom_histogram()
```



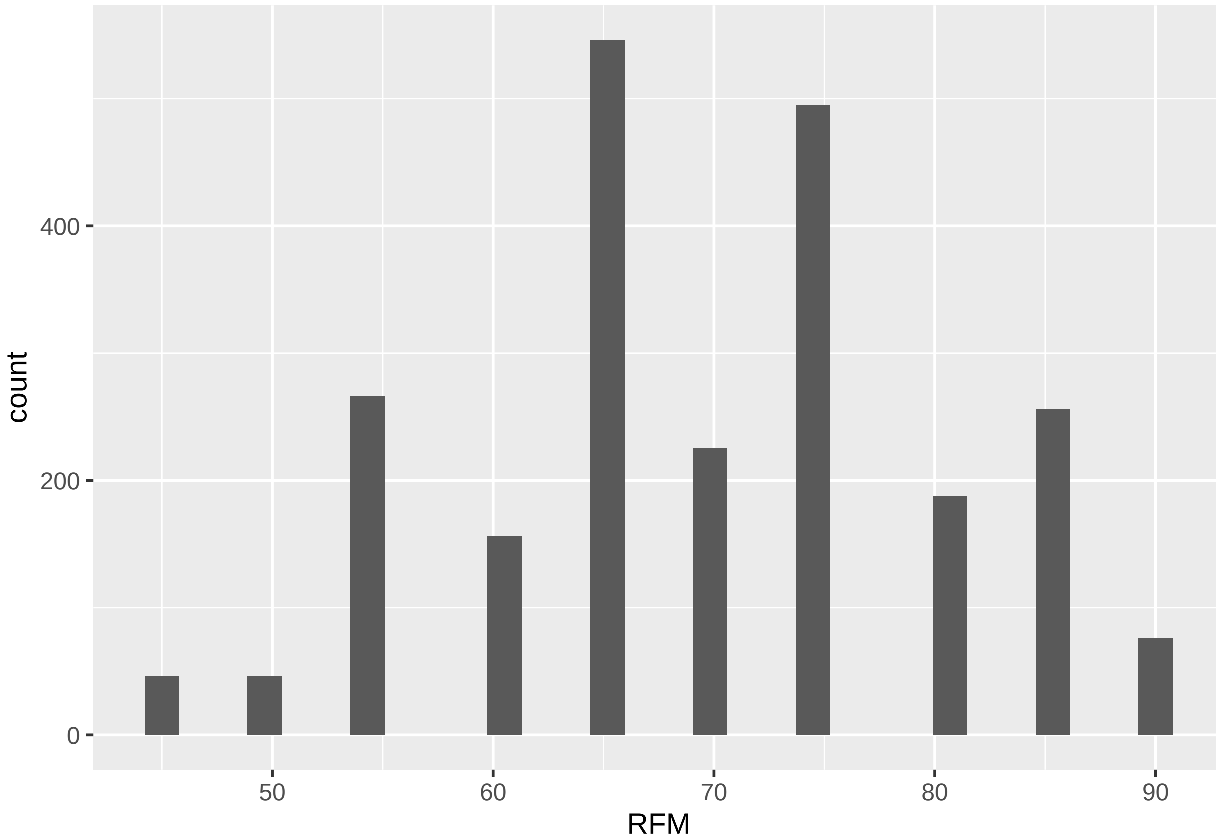

```
ggplot(BBBClub.holdout, aes(F)) +
  geom_histogram()
```




```
ggplot(BBBClub.holdout, aes(M)) +
  geom_histogram()
```




```
ggplot(BBBClub.holdout, aes(RFM)) +
  geom_histogram()
```



The linear model

Now we calibrate the model using a Linear regression model. We use the choice data set (historic data)

```
BBBClub.lm <- lm(choice~gender+amount+freq+last+first+child+youth+cook+diy+art, data=BBBClub.choice)

#install.packages("stargazer")
library(stargazer)
stargazer(BBBClub.lm, type="text")
```

```
##
## ===============================================
##                          Dependent variable:
##                      ---------------------------
##                                choice
## -----------------------------------------------
## gender                        -0.131***
##                                (0.020)
##
## amount                        0.0003**
##                               (0.0001)
##
## freq                          -0.009***
##                                (0.002)
```

```
##
## last                         0.097***
##                               (0.014)
##
## first                         -0.002
##                               (0.002)
##
## child                        -0.126***
##                               (0.016)
##
## youth                        -0.096***
##                               (0.020)
##
## cook                         -0.141***
##                               (0.017)
##
## diy                          -0.135***
##                               (0.020)
##
## art                          0.118***
##                               (0.019)
##
## Constant                     0.364***
##                               (0.031)
##
## -----------------------------------------------
## Observations                   1,600
## R2                             0.240
## Adjusted R2                    0.235
## Residual Std. Error      0.379 (df = 1589)
## F Statistic          50.196*** (df = 10; 1589)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

Use the calibrated model to predict choice on the holdout sample (we predict the probability of buying the art book using de model from choice data)

```
BBBClub.holdout$plm <- predict(BBBClub.lm, BBBClub.holdout)
knitr::kable(
  t(head(BBBClub.holdout)),
  digits=2,
  caption = "Descripción de los 6 primeras individuos de los nuevos datos")
```

Table 4: Descripción de los 6 primeras individuos de los nuevos datos

	1	2	3	4	5	6
choice	1.0	1.00	1.0	1.0	1.00	1.00
gender	0.0	1.00	1.0	0.0	1.00	0.00
amount	287.0	215.00	261.0	24.0	120.00	66.00
freq	12.0	4.00	2.0	4.0	8.00	2.00
last	4.0	1.00	1.0	1.0	1.00	4.00

	1	2	3	4	5	6
first	24.0	4.00	2.0	4.0	8.00	16.00
child	0.0	0.00	0.0	1.0	0.00	0.00
youth	3.0	0.00	0.0	0.0	0.00	0.00
cook	0.0	0.00	0.0	0.0	0.00	1.00
diy	0.0	0.00	0.0	0.0	0.00	1.00
art	1.0	1.00	1.0	0.0	1.00	1.00
R	20.0	25.00	25.0	25.0	25.00	20.00
F	20.0	10.00	10.0	10.0	10.00	10.00
M	40.0	30.00	40.0	10.0	20.00	20.00
RFM	80.0	65.00	75.0	45.0	55.00	50.00
plm	0.5	0.46	0.5	0.3	0.39	0.56

The logistic model

Use the logistic model with the holdout sample (we predict the probability of buying the art book using de model from choice data using a Logistic regression model)

```
BBBClub.glm<-glm(
  choice~gender+amount+freq+last+first+child+youth+cook+diy+art,
  family=binomial(),
  data=BBBClub.choice)
stargazer(BBBClub.glm, type="text" )
```

```
##
## =============================================
##                       Dependent variable:
##                   ---------------------------
##                             choice
## ---------------------------------------------
## gender                    -0.863***
##                            (0.137)
##
## amount                     0.002**
##                            (0.001)
##
## freq                      -0.076***
##                            (0.017)
##
## last                      0.612***
##                            (0.094)
##
## first                      -0.015
##                            (0.013)
##
## child                     -0.811***
##                            (0.117)
##
## youth                     -0.637***
##                            (0.143)
##
```

```
## cook                          -0.923***
##                                (0.119)
##
## diy                           -0.906***
##                                (0.144)
##
## art                           0.686***
##                                (0.127)
##
## Constant                       -0.352
##                                (0.214)
##
## ---------------------------------------------
## Observations                    1,600
## Log Likelihood                -696.080
## Akaike Inf. Crit.             1,414.159
## =============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

```
BBBClub.holdout$pglm <- predict(
  BBBClub.glm,
  BBBClub.holdout,
  type="response")
```

Test accuracy of prediction

```
confusion.glm(BBBClub.choice, BBBClub.glm)
```

```
##       FALSE TRUE class.error
## FALSE  1120  240  0.06666667
## TRUE     80  160  0.60000000
```

See the new set of variables

```
knitr::kable(
  t(head(BBBClub.holdout)),
  digits=2,
  caption = "Descripción de los 6 primeras individuos de los nuevos datos")
```

Table 5: Descripción de los 6 primeras individuos de los nuevos datos

	1	2	3	4	5	6
choice	1.00	1.00	1.0	1.0	1.00	1.00
gender	0.00	1.00	1.0	0.0	1.00	0.00
amount	287.00	215.00	261.0	24.0	120.00	66.00
freq	12.00	4.00	2.0	4.0	8.00	2.00
last	4.00	1.00	1.0	1.0	1.00	4.00

	1	2	3	4	5	6
first	24.00	4.00	2.0	4.0	8.00	16.00
child	0.00	0.00	0.0	1.0	0.00	0.00
youth	3.00	0.00	0.0	0.0	0.00	0.00
cook	0.00	0.00	0.0	0.0	0.00	1.00
diy	0.00	0.00	0.0	0.0	0.00	1.00
art	1.00	1.00	1.0	0.0	1.00	1.00
R	20.00	25.00	25.0	25.0	25.00	20.00
F	20.00	10.00	10.0	10.0	10.00	10.00
M	40.00	30.00	40.0	10.0	20.00	20.00
RFM	80.00	65.00	75.0	45.0	55.00	50.00
plm	0.50	0.46	0.5	0.3	0.39	0.56
pglm	0.54	0.53	0.6	0.3	0.40	0.67

Apply the calibrated models to make marketing decisions: select the segments that will be the target of our marketing efforts

Divide individuals in 10 segments according to quantiles

```
quantile(BBBClub.holdout$RFM, probs=seq(.1,1,.1))
```

```
##  10%  20%  30%  40%  50%  60%  70%  80%  90% 100% 
##   55   60   65   65   70   75   75   80   85   90
```

```
quantile(BBBClub.holdout$plm, probs=seq(.1,1,.1))
```

```
##          10%          20%          30%          40%          50%          60% 
## -0.003423609  0.060944951  0.105823254  0.148539537  0.190511435  0.228692883 
##          70%          80%          90%         100% 
##  0.279312194  0.347923157  0.451071244  1.035033938
```

Use the models por selecting the target of our marketing efforts

Number of people that bought the book in the choice sample

```
exit<-sum(BBBClub.holdout$choice)
exit
```

```
## [1] 204
```

Using the logistic model

```r
library(tidyverse)
BBBClub.holdout.bypglm<-arrange(BBBClub.holdout, desc(pglm))

knitr::kable(
  t(head(BBBClub.holdout.bypglm)),
  digits=2,
  caption = "Descripción de los 6 primeras individuos de los nuevos datos")
```

Table 6: Descripción de los 6 primeras individuos de los nuevos datos

	1	2	3	4	5	6
choice	0.00	1.00	1.00	1.00	1.00	0.00
gender	1.00	1.00	0.00	0.00	0.00	0.00
amount	221.00	250.00	201.00	309.00	225.00	371.00
freq	10.00	20.00	4.00	6.00	2.00	10.00
last	7.00	12.00	7.00	4.00	2.00	7.00
first	36.00	88.00	22.00	18.00	8.00	28.00
child	0.00	1.00	1.00	0.00	0.00	2.00
youth	0.00	0.00	1.00	0.00	0.00	0.00
cook	0.00	2.00	0.00	0.00	0.00	1.00
diy	1.00	1.00	1.00	1.00	0.00	0.00
art	3.00	4.00	2.00	2.00	2.00	2.00
R	20.00	20.00	20.00	20.00	25.00	20.00
F	20.00	20.00	10.00	10.00	10.00	20.00
M	30.00	30.00	30.00	40.00	30.00	50.00
RFM	70.00	70.00	60.00	70.00	65.00	90.00
plm	1.03	1.04	0.90	0.85	0.82	0.84
pglm	0.97	0.95	0.94	0.92	0.92	0.91

```r
#Prepare the report for logistic model
source("marketing-models.R")
```

save quartiles

The `ntile` function creates groups according to a variable. In this case the variable is `BBBClub.holdout.bypglm$pglm` and will form 10 groups. Using the argument `reverse=TRUE` it would be necessary the function `reverse.quartile`.

The function `reverse.quartile` in the `marketing-models.R` script changes the order or quartiles, so that the ones with the highest probability are placed in the first decile, and adds to the data base a new variable call `decil`.

```r
BBBClub.holdout.bypglm$quartile <- ntile(BBBClub.holdout.bypglm$pglm, 10)

BBBClub.holdout.bypglm <- reverse.quartile(BBBClub.holdout.bypglm)
```

check

```r
knitr::kable(t(head(BBBClub.holdout.bypglm)), digits=2, caption = "Descripción de los 6 primeras individ
```

Table 7: Descripción de los 6 primeras individuos de los nuevos datos

	1	2	3	4	5	6
choice	0.00	1.00	1.00	1.00	1.00	0.00
gender	1.00	1.00	0.00	0.00	0.00	0.00
amount	221.00	250.00	201.00	309.00	225.00	371.00
freq	10.00	20.00	4.00	6.00	2.00	10.00
last	7.00	12.00	7.00	4.00	2.00	7.00
first	36.00	88.00	22.00	18.00	8.00	28.00
child	0.00	1.00	1.00	0.00	0.00	2.00
youth	0.00	0.00	1.00	0.00	0.00	0.00
cook	0.00	2.00	0.00	0.00	0.00	1.00
diy	1.00	1.00	1.00	1.00	0.00	0.00
art	3.00	4.00	2.00	2.00	2.00	2.00
R	20.00	20.00	20.00	20.00	25.00	20.00
F	20.00	20.00	10.00	10.00	10.00	20.00
M	30.00	30.00	30.00	40.00	30.00	50.00
RFM	70.00	70.00	60.00	70.00	65.00	90.00
plm	1.03	1.04	0.90	0.85	0.82	0.84
pglm	0.97	0.95	0.94	0.92	0.92	0.91
quartile	10.00	10.00	10.00	10.00	10.00	10.00
decil	1.00	1.00	1.00	1.00	1.00	1.00

Cummulative distribution of success (what would have happened if we had used glm to target segments)

```r
BBBClub.holdout.bypglm$choice2<-BBBClub.holdout.bypglm$choice/sum(BBBClub.holdout.bypglm$choice)

BBBClub.holdout.bypglm$acumul<-cumsum(BBBClub.holdout.bypglm$choice2)

knitr::kable(t(head(BBBClub.holdout.bypglm)), digits=2, caption = "Descripción de los 6 primeras individ
```

Table 8: Descripción de los 6 primeras individuos de los nuevos datos

	1	2	3	4	5	6
choice	0.00	1.00	1.00	1.00	1.00	0.00
gender	1.00	1.00	0.00	0.00	0.00	0.00
amount	221.00	250.00	201.00	309.00	225.00	371.00
freq	10.00	20.00	4.00	6.00	2.00	10.00
last	7.00	12.00	7.00	4.00	2.00	7.00
first	36.00	88.00	22.00	18.00	8.00	28.00
child	0.00	1.00	1.00	0.00	0.00	2.00
youth	0.00	0.00	1.00	0.00	0.00	0.00
cook	0.00	2.00	0.00	0.00	0.00	1.00

	1	2	3	4	5	6
diy	1.00	1.00	1.00	1.00	0.00	0.00
art	3.00	4.00	2.00	2.00	2.00	2.00
R	20.00	20.00	20.00	20.00	25.00	20.00
F	20.00	20.00	10.00	10.00	10.00	20.00
M	30.00	30.00	30.00	40.00	30.00	50.00
RFM	70.00	70.00	60.00	70.00	65.00	90.00
plm	1.03	1.04	0.90	0.85	0.82	0.84
pglm	0.97	0.95	0.94	0.92	0.92	0.91
quartile	10.00	10.00	10.00	10.00	10.00	10.00
decil	1.00	1.00	1.00	1.00	1.00	1.00
choice2	0.00	0.00	0.00	0.00	0.00	0.00
acumul	0.00	0.00	0.01	0.01	0.02	0.02

by quartile count the success ratio

```
table1<-BBBClub.holdout.bypglm %>%
  group_by(decil) %>%
  summarize(
    count=n(),
    mean.choice=sum(choice)/204
  )
table1
```

```
## # A tibble: 10 x 3
##    decil count mean.choice
##    <dbl> <int>       <dbl>
##  1     1   230     0.422
##  2     2   230     0.167
##  3     3   230     0.123
##  4     4   230     0.0882
##  5     5   230     0.0539
##  6     6   230     0.0294
##  7     7   230     0.0588
##  8     8   230     0.0343
##  9     9   230     0.0196
## 10    10   230     0.00490
```

Add the cumulative distribution

```
table1$acumul<-cumsum(table1$mean.choice)
table1
```

```
## # A tibble: 10 x 4
##    decil count mean.choice acumul
##    <dbl> <int>       <dbl>  <dbl>
##  1     1   230     0.422    0.422
##  2     2   230     0.167    0.588
```

```
## 3      3   230      0.123     0.711
## 4      4   230      0.0882    0.799
## 5      5   230      0.0539    0.853
## 6      6   230      0.0294    0.882
## 7      7   230      0.0588    0.941
## 8      8   230      0.0343    0.975
## 9      9   230      0.0196    0.995
## 10    10   230      0.00490   1
```

Add the cumulative mailing to be sent

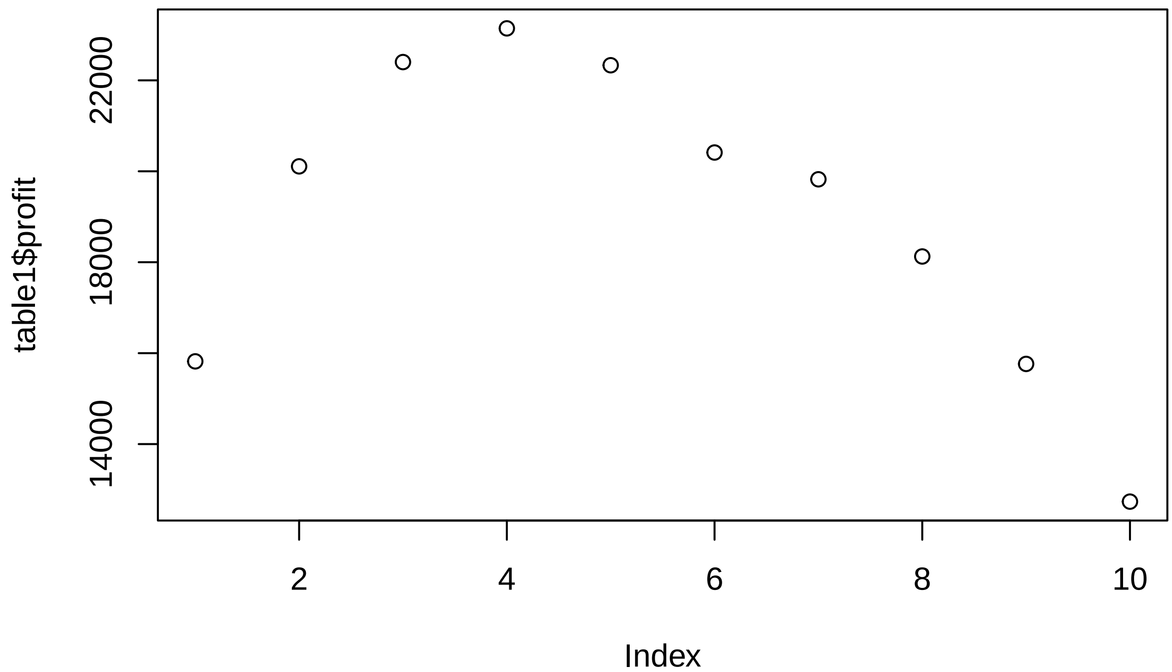
```
units.mailed<- c(5000,10000,15000,20000,25000,30000,35000,40000,45000,50000)
#units.mailed
#add them to table
table1$mailed<-units.mailed
```

Now we add the cost, the market potencial, the units sold, the margin and profit

```
#Add cost
table1$cost<-table1$mailed*0.65
#Add market potential
market.potential<-50000*(exit/2300)
#Add units sold
table1$sold<-table1$acumul*(market.potential)
#table1
#Add profit
margin<-31.95-1.45*15
#margin
table1$profit<-table1$sold*margin-table1$cost
table1
```

```
## # A tibble: 10 x 8
##    decil count mean.choice acumul mailed  cost  sold profit
##    <dbl> <int>       <dbl>  <dbl>  <dbl> <dbl> <dbl>  <dbl>
## 1      1   230     0.422    0.422   5000  3250 1870. 15820.
## 2      2   230     0.167    0.588  10000  6500 2609. 20109.
## 3      3   230     0.123    0.711  15000  9750 3152. 22402.
## 4      4   230     0.0882   0.799  20000 13000 3543. 23143.
## 5      5   230     0.0539   0.853  25000 16250 3783. 22333.
## 6      6   230     0.0294   0.882  30000 19500 3913. 20413.
## 7      7   230     0.0588   0.941  35000 22750 4174. 19824.
## 8      8   230     0.0343   0.975  40000 26000 4326. 18126.
## 9      9   230     0.0196   0.995  45000 29250 4413. 15763.
## 10    10   230     0.00490  1      50000 32500 4435. 12735.
```

```
plot(table1$profit)
```



Check if we improve the profit seleccting onlya few segments

```
print("Total Income")
```

```
## [1] "Total Income"
```

```
sum(table1$profit)
```

```
## [1] 190667.4
```

```
print("Total Costs")
```

```
## [1] "Total Costs"
```

```
sum(table1$cost)
```

```
## [1] 178750
```

```
print("Total Profit")
```

```
## [1] "Total Profit"
```

```
total.profit1<-sum(table1$profit) - sum(table1$cost)
total.profit1
```

```
## [1] 11917.39
```

```
print("Total relative profit")
```

```
## [1] "Total relative profit"
```

```
rendimiento1<-total.profit1/sum(table1$cost)
rendimiento1
```

```
## [1] 0.06667072
```

Use linear model for selecting the segments to be the target of marketing efforts

order the data base for linear model

```
BBBClub.holdout.byplm<-arrange(BBBClub.holdout, desc(plm))
knitr::kable(t(head(BBBClub.holdout.byplm)), digits=2, caption = "Descripción de los 6 primeras individu
```

Table 9: Descripción de los 6 primeras individuos de los nuevos datos

	1	2	3	4	5	6
choice	1.00	0.00	1.00	1.00	1.00	0.00
gender	1.00	1.00	0.00	0.00	0.00	0.00
amount	250.00	221.00	201.00	444.00	309.00	371.00
freq	20.00	10.00	4.00	10.00	6.00	10.00
last	12.00	7.00	7.00	11.00	4.00	7.00
first	88.00	36.00	22.00	66.00	18.00	28.00
child	1.00	0.00	1.00	1.00	0.00	2.00
youth	0.00	0.00	1.00	0.00	0.00	0.00
cook	2.00	0.00	0.00	3.00	0.00	1.00
diy	1.00	1.00	1.00	2.00	1.00	0.00
art	4.00	3.00	2.00	3.00	2.00	2.00
R	20.00	20.00	20.00	20.00	20.00	20.00
F	20.00	20.00	10.00	20.00	10.00	20.00
M	30.00	30.00	30.00	50.00	40.00	50.00
RFM	70.00	70.00	60.00	90.00	70.00	90.00
plm	1.04	1.03	0.90	0.86	0.85	0.84
pglm	0.95	0.97	0.94	0.89	0.92	0.91

Prepare the report for linear model

```r
source("marketing-models.R")

#save quartiles
BBBClub.holdout.byplm$quartile <- ntile(BBBClub.holdout.byplm$plm, 10)
BBBClub.holdout.byplm <- reverse.quartile(BBBClub.holdout.byplm)
#check
knitr::kable(t(head(BBBClub.holdout.byplm)), digits=2, caption = "Descripción de los 6 primeras individu
```

Table 10: Descripción de los 6 primeras individuos de los nuevos datos

	1	2	3	4	5	6
choice	1.00	0.00	1.00	1.00	1.00	0.00
gender	1.00	1.00	0.00	0.00	0.00	0.00
amount	250.00	221.00	201.00	444.00	309.00	371.00
freq	20.00	10.00	4.00	10.00	6.00	10.00
last	12.00	7.00	7.00	11.00	4.00	7.00
first	88.00	36.00	22.00	66.00	18.00	28.00
child	1.00	0.00	1.00	1.00	0.00	2.00
youth	0.00	0.00	1.00	0.00	0.00	0.00
cook	2.00	0.00	0.00	3.00	0.00	1.00
diy	1.00	1.00	1.00	2.00	1.00	0.00
art	4.00	3.00	2.00	3.00	2.00	2.00
R	20.00	20.00	20.00	20.00	20.00	20.00
F	20.00	20.00	10.00	20.00	10.00	20.00
M	30.00	30.00	30.00	50.00	40.00	50.00
RFM	70.00	70.00	60.00	90.00	70.00	90.00
plm	1.04	1.03	0.90	0.86	0.85	0.84
pglm	0.95	0.97	0.94	0.89	0.92	0.91
quartile	10.00	10.00	10.00	10.00	10.00	10.00
decil	1.00	1.00	1.00	1.00	1.00	1.00

Cummulative probability of success

```r
BBBClub.holdout.byplm$choice2<-BBBClub.holdout.byplm$choice/sum(BBBClub.holdout.byplm$choice)
BBBClub.holdout.byplm$acumul<-cumsum(BBBClub.holdout.byplm$choice2)
```

Prepare the table for seleccting the segments

```r
#by quartile count the exit ratio

table.lm<-BBBClub.holdout.byplm %>%
  group_by(decil) %>%
  summarize(
    count=n(),
    mean.choice=sum(choice)/204
  )
table.lm
```

```
## # A tibble: 10 x 3
##    decil count mean.choice
##    <dbl> <int>       <dbl>
##  1     1   230     0.422
##  2     2   230     0.167
##  3     3   230     0.118
##  4     4   230     0.0784
##  5     5   230     0.0686
##  6     6   230     0.0588
##  7     7   230     0.0294
##  8     8   230     0.0343
##  9     9   230     0.0196
## 10    10   230     0.00490
```

Add the cumulative distribution

```
table.lm$acumul<-cumsum(table.lm$mean.choice)
table.lm
```

```
## # A tibble: 10 x 4
##    decil count mean.choice acumul
##    <dbl> <int>       <dbl>  <dbl>
##  1     1   230     0.422    0.422
##  2     2   230     0.167    0.588
##  3     3   230     0.118    0.706
##  4     4   230     0.0784   0.784
##  5     5   230     0.0686   0.853
##  6     6   230     0.0588   0.912
##  7     7   230     0.0294   0.941
##  8     8   230     0.0343   0.975
##  9     9   230     0.0196   0.995
## 10    10   230     0.00490  1
```

Add the cumulative mailing to be sent and add it to the table

```
units.mailed<- c(5000,10000,15000,20000,25000,30000,35000,40000,45000,50000)
units.mailed
```

```
##  [1]  5000 10000 15000 20000 25000 30000 35000 40000 45000 50000
```

```
#add them to table
table.lm$mailed<-units.mailed
```

Add cost, market potential, units sold, margin and profit tothe table

```
table.lm$cost<-table.lm$mailed*0.65
#Add market potential
```

```r
market.potential<-50000*(exit/2300)
#Add units sold
table.lm$sold<-table.lm$acumul*(market.potential)
table.lm
```

```
## # A tibble: 10 x 7
##    decil count mean.choice acumul mailed  cost  sold
##    <dbl> <int>       <dbl>  <dbl>  <dbl> <dbl> <dbl>
##  1     1   230     0.422    0.422   5000  3250 1870.
##  2     2   230     0.167    0.588  10000  6500 2609.
##  3     3   230     0.118    0.706  15000  9750 3130.
##  4     4   230     0.0784   0.784  20000 13000 3478.
##  5     5   230     0.0686   0.853  25000 16250 3783.
##  6     6   230     0.0588   0.912  30000 19500 4043.
##  7     7   230     0.0294   0.941  35000 22750 4174.
##  8     8   230     0.0343   0.975  40000 26000 4326.
##  9     9   230     0.0196   0.995  45000 29250 4413.
## 10    10   230     0.00490  1      50000 32500 4435.
```

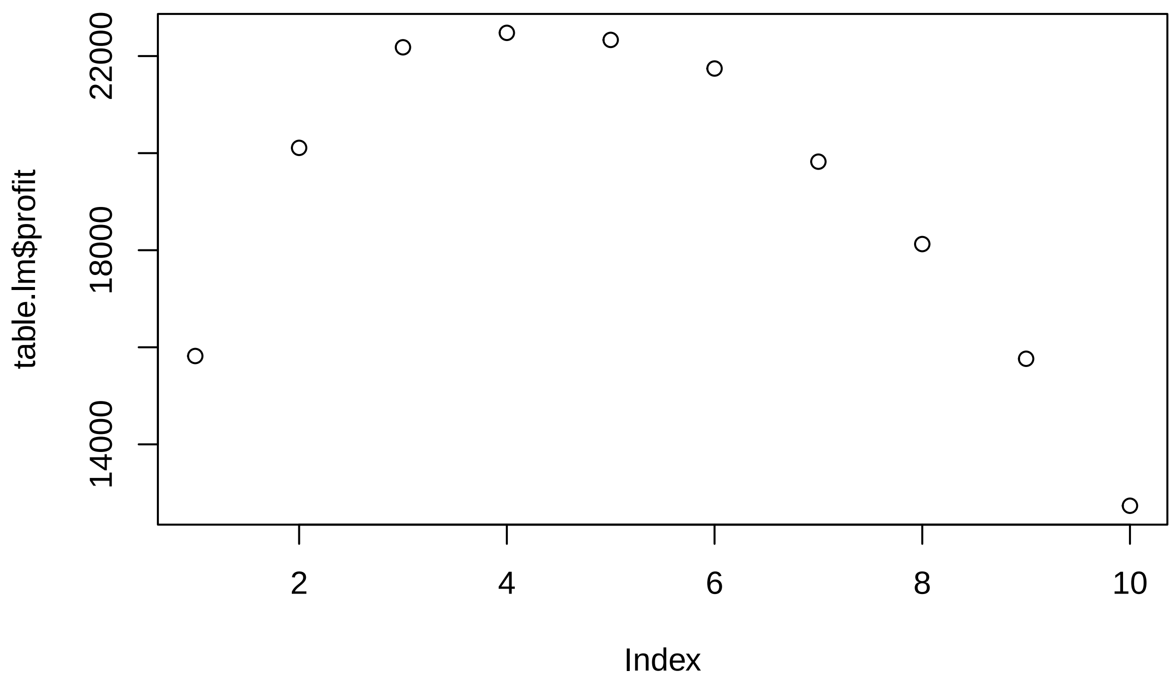
```r
#Add profit
margin<-31.95-1.45*15
margin
```

```
## [1] 10.2
```

```r
table.lm$profit<-table.lm$sold*margin-table.lm$cost
table.lm
```

```
## # A tibble: 10 x 8
##    decil count mean.choice acumul mailed  cost  sold profit
##    <dbl> <int>       <dbl>  <dbl>  <dbl> <dbl> <dbl>  <dbl>
##  1     1   230     0.422    0.422   5000  3250 1870. 15820.
##  2     2   230     0.167    0.588  10000  6500 2609. 20109.
##  3     3   230     0.118    0.706  15000  9750 3130. 22180.
##  4     4   230     0.0784   0.784  20000 13000 3478. 22478.
##  5     5   230     0.0686   0.853  25000 16250 3783. 22333.
##  6     6   230     0.0588   0.912  30000 19500 4043. 21743.
##  7     7   230     0.0294   0.941  35000 22750 4174. 19824.
##  8     8   230     0.0343   0.975  40000 26000 4326. 18126.
##  9     9   230     0.0196   0.995  45000 29250 4413. 15763.
## 10    10   230     0.00490  1      50000 32500 4435. 12735.
```

```r
plot(table.lm$profit)
```



Check if we improve the profit seleccting onlya few segments

```
print("Total Income")
```

```
## [1] "Total Income"
```

```
sum(table.lm$profit)
```

```
## [1] 191110.9
```

```
print("Total cost")
```

```
## [1] "Total cost"
```

```
sum(table.lm$cost)
```

```
## [1] 178750
```

```
total.profit.lm<-sum(table.lm$profit) - sum(table.lm$cost)
print("Total profit")
```

```
## [1] "Total profit"
```

```r
total.profit.lm
```

```
## [1] 12360.87
```

```r
rendimiento.lm<-total.profit.lm/sum(table.lm$cost)
print("Relative profit")
```

```
## [1] "Relative profit"
```

```r
rendimiento.lm
```

```
## [1] 0.06915172
```

RFM model report

order the data base for RFM model

```r
BBBClub.holdout.byRFM<-arrange(BBBClub.holdout, desc(RFM))

knitr::kable(t(head(BBBClub.holdout.byRFM)), digits=2, caption = "Descripción de los 6 primeras individu
```

Table 11: Descripción de los 6 primeras individuos de los nuevos datos

	1	2	3	4	5	6
choice	1.00	1.0	1.00	1.00	1.00	1.00
gender	1.00	0.0	0.00	0.00	0.00	0.00
amount	377.00	400.0	410.00	360.00	444.00	427.00
freq	16.00	10.0	10.00	10.00	10.00	14.00
last	9.00	7.0	11.00	9.00	11.00	11.00
first	44.00	32.0	68.00	60.00	66.00	74.00
child	3.00	0.0	3.00	3.00	1.00	4.00
youth	0.00	0.0	1.00	1.00	0.00	1.00
cook	2.00	2.0	1.00	0.00	3.00	1.00
diy	0.00	1.0	3.00	2.00	2.00	1.00
art	3.00	1.0	1.00	3.00	3.00	2.00
R	20.00	20.0	20.00	20.00	20.00	20.00
F	20.00	20.0	20.00	20.00	20.00	20.00
M	50.00	50.0	50.00	50.00	50.00	50.00
RFM	90.00	90.0	90.00	90.00	90.00	90.00
plm	0.67	0.7	0.41	0.73	0.86	0.63
pglm	0.71	0.8	0.34	0.80	0.89	0.66

```r
#Prepare the report for logistic model
source("marketing-models.R")
```

Devide data according to quartiles

```
#save quartiles
BBBClub.holdout.byRFM$quartile <- ntile(BBBClub.holdout.byRFM$RFM, 10)
BBBClub.holdout.byRFM <- reverse.quartile(BBBClub.holdout.byRFM)

#check
knitr::kable(t(head(BBBClub.holdout.byRFM)), digits=2, caption = "Descripción de los 6 primeras individu
```

Table 12: Descripción de los 6 primeras individuos de los nuevos datos

	1	2	3	4	5	6
choice	1.00	1.0	1.00	1.00	1.00	1.00
gender	1.00	0.0	0.00	0.00	0.00	0.00
amount	377.00	400.0	410.00	360.00	444.00	427.00
freq	16.00	10.0	10.00	10.00	10.00	14.00
last	9.00	7.0	11.00	9.00	11.00	11.00
first	44.00	32.0	68.00	60.00	66.00	74.00
child	3.00	0.0	3.00	3.00	1.00	4.00
youth	0.00	0.0	1.00	1.00	0.00	1.00
cook	2.00	2.0	1.00	0.00	3.00	1.00
diy	0.00	1.0	3.00	2.00	2.00	1.00
art	3.00	1.0	1.00	3.00	3.00	2.00
R	20.00	20.0	20.00	20.00	20.00	20.00
F	20.00	20.0	20.00	20.00	20.00	20.00
M	50.00	50.0	50.00	50.00	50.00	50.00
RFM	90.00	90.0	90.00	90.00	90.00	90.00
plm	0.67	0.7	0.41	0.73	0.86	0.63
pglm	0.71	0.8	0.34	0.80	0.89	0.66
quartile	10.00	10.0	10.00	10.00	10.00	10.00
decil	1.00	1.0	1.00	1.00	1.00	1.00

Cummulative distribution of success

```
BBBClub.holdout.byRFM$choice2<-BBBClub.holdout.byRFM$choice/sum(BBBClub.holdout.byRFM$choice)
BBBClub.holdout.byRFM$acumul<-cumsum(BBBClub.holdout.byRFM$choice2)
```

Prepare the summary table

```
#by quartile count the exit ratio

table.RFM<-BBBClub.holdout.byRFM %>%
  group_by(decil) %>%
  summarize(
    count=n(),
    mean.choice=sum(choice)/204,
    mean.RFM=mean(RFM)
```

```r
  )
table.RFM
```

```
## # A tibble: 10 x 4
##    decil count mean.choice mean.RFM
##    <dbl> <int>       <dbl>    <dbl>
##  1     1   230      0.0343     86.7
##  2     2   230      0.103      82.2
##  3     3   230      0.0637     76.3
##  4     4   230      0          75
##  5     5   230      0.211      72.1
##  6     6   230      0.147      67.0
##  7     7   230      0          65
##  8     8   230      0.176      63.8
##  9     9   230      0.0980     57.2
## 10    10   230      0.167      52
```

Add the cumulative distribution

```r
#Add the cumulative distribution
table.RFM$acumul<-cumsum(table.RFM$mean.choice)
table.RFM
```

```
## # A tibble: 10 x 5
##    decil count mean.choice mean.RFM acumul
##    <dbl> <int>       <dbl>    <dbl>  <dbl>
##  1     1   230      0.0343     86.7 0.0343
##  2     2   230      0.103      82.2 0.137
##  3     3   230      0.0637     76.3 0.201
##  4     4   230      0          75   0.201
##  5     5   230      0.211      72.1 0.412
##  6     6   230      0.147      67.0 0.559
##  7     7   230      0          65   0.559
##  8     8   230      0.176      63.8 0.735
##  9     9   230      0.0980     57.2 0.833
## 10    10   230      0.167      52   1
```

```r
#Add the cumulative mailing to be sent
units.mailed<- c(5000,10000,15000,20000,25000,30000,35000,40000,45000,50000)
#units.mailed
#add them to table
table.RFM$mailed<-units.mailed
table.RFM
```

```
## # A tibble: 10 x 6
##    decil count mean.choice mean.RFM acumul mailed
##    <dbl> <int>       <dbl>    <dbl>  <dbl>  <dbl>
##  1     1   230      0.0343     86.7 0.0343   5000
##  2     2   230      0.103      82.2 0.137   10000
##  3     3   230      0.0637     76.3 0.201   15000
##  4     4   230      0          75   0.201   20000
```

```
## 5      5    230       0.211       72.1 0.412    25000
## 6      6    230       0.147       67.0 0.559    30000
## 7      7    230       0          65    0.559    35000
## 8      8    230       0.176       63.8 0.735    40000
## 9      9    230       0.0980      57.2 0.833    45000
## 10    10    230       0.167       52    1        50000
```

Add cost, units sold, margin, and profit

```
#Add cost
table.RFM$cost<-table.RFM$mailed*0.65
#Add market potential
market.potential<-50000*(exit/2300)
#Add units sold
table.RFM$sold<-table.RFM$acumul*(market.potential)
table.RFM
```

```
## # A tibble: 10 x 8
##    decil count mean.choice mean.RFM acumul mailed  cost  sold
##    <dbl> <int>       <dbl>    <dbl>  <dbl>  <dbl> <dbl> <dbl>
##  1     1   230      0.0343     86.7 0.0343   5000  3250  152.
##  2     2   230      0.103      82.2 0.137   10000  6500  609.
##  3     3   230      0.0637     76.3 0.201   15000  9750  891.
##  4     4   230      0          75   0.201   20000 13000  891.
##  5     5   230      0.211      72.1 0.412   25000 16250 1826.
##  6     6   230      0.147      67.0 0.559   30000 19500 2478.
##  7     7   230      0          65   0.559   35000 22750 2478.
##  8     8   230      0.176      63.8 0.735   40000 26000 3261.
##  9     9   230      0.0980     57.2 0.833   45000 29250 3696.
## 10    10   230      0.167      52   1       50000 32500 4435.
```

```
#Add profit
margin<-31.95-1.45*15
margin
```

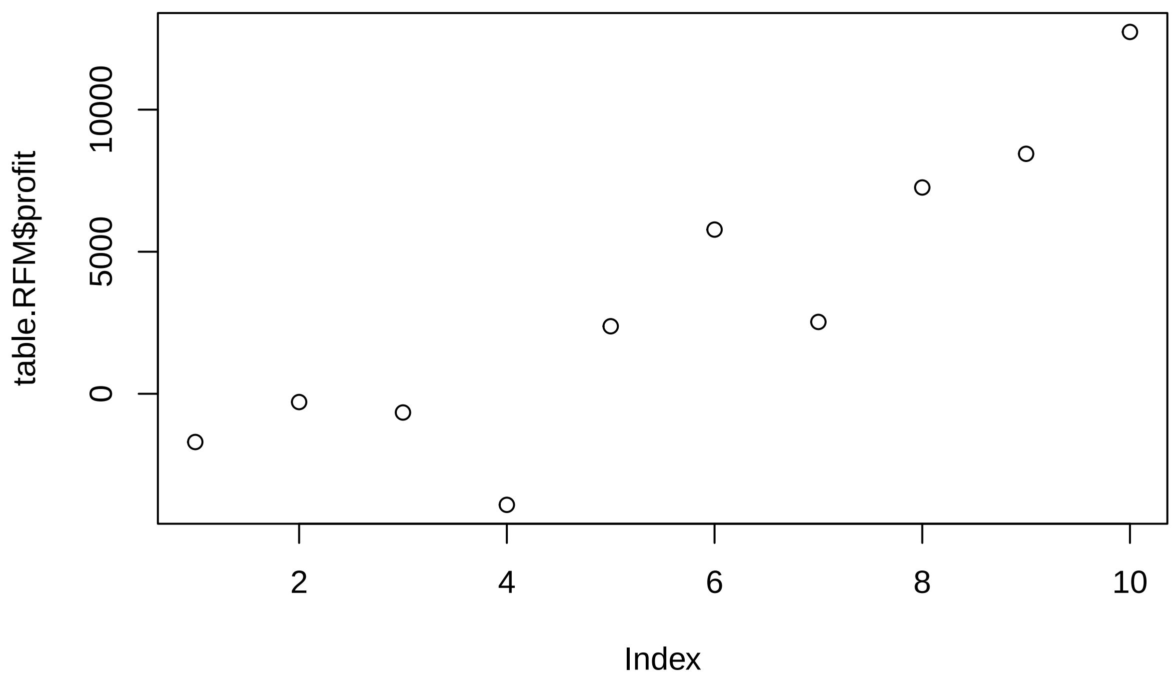
```
## [1] 10.2
```

```
table.RFM$profit<-table.RFM$sold*margin-table.RFM$cost
table.RFM
```

```
## # A tibble: 10 x 9
##    decil count mean.choice mean.RFM acumul mailed  cost  sold profit
##    <dbl> <int>       <dbl>    <dbl>  <dbl>  <dbl> <dbl> <dbl>  <dbl>
##  1     1   230      0.0343     86.7 0.0343   5000  3250  152. -1698.
##  2     2   230      0.103      82.2 0.137   10000  6500  609.  -291.
##  3     3   230      0.0637     76.3 0.201   15000  9750  891.  -659.
##  4     4   230      0          75   0.201   20000 13000  891. -3909.
##  5     5   230      0.211      72.1 0.412   25000 16250 1826.  2376.
##  6     6   230      0.147      67.0 0.559   30000 19500 2478.  5778.
##  7     7   230      0          65   0.559   35000 22750 2478.  2528.
##  8     8   230      0.176      63.8 0.735   40000 26000 3261.  7261.
```

```
## 9       9    230       0.0980      57.2 0.833    45000 29250 3696.  8446.
## 10     10    230       0.167       52   1        50000 32500 4435. 12735.
```

```
plot(table.RFM$profit)
```



Compare results

```
print("Total income:")
```

```
## [1] "Total income:"
```

```
sum(table.RFM$profit)
```

```
## [1] 32567.39
```

```
print("Total cost:")
```

```
## [1] "Total cost:"
```

```
sum(table.RFM$cost)
```

```
## [1] 178750
```

```
print("Total profit:")
```

```
## [1] "Total profit:"
```

```
total.profit.RFM<-sum(table.RFM$profit) - sum(table.RFM$cost)
total.profit.RFM
```

```
## [1] -146182.6
```

```
print("Total relative profit:")
```

```
## [1] "Total relative profit:"
```

```
rendimiento.RFM<-total.profit.RFM/sum(table.RFM$cost)
rendimiento.RFM
```

```
## [1] -0.8178048
```

Compare findings

The glm model

```
knitr::kable(table1, digits=2, caption = "Logistic model")
```

Table 13: Logistic model

decil	count	mean.choice	acumul	mailed	cost	sold	profit
1	230	0.42	0.42	5000	3250	1869.57	15819.57
2	230	0.17	0.59	10000	6500	2608.70	20108.70
3	230	0.12	0.71	15000	9750	3152.17	22402.17
4	230	0.09	0.80	20000	13000	3543.48	23143.48
5	230	0.05	0.85	25000	16250	3782.61	22332.61
6	230	0.03	0.88	30000	19500	3913.04	20413.04
7	230	0.06	0.94	35000	22750	4173.91	19823.91
8	230	0.03	0.98	40000	26000	4326.09	18126.09
9	230	0.02	1.00	45000	29250	4413.04	15763.04
10	230	0.00	1.00	50000	32500	4434.78	12734.78

```
knitr::kable(table.lm, digits=2, caption = "Linear model")
```

Table 14: Linear model

decil	count	mean.choice	acumul	mailed	cost	sold	profit
1	230	0.42	0.42	5000	3250	1869.57	15819.57
2	230	0.17	0.59	10000	6500	2608.70	20108.70
3	230	0.12	0.71	15000	9750	3130.43	22180.43
4	230	0.08	0.78	20000	13000	3478.26	22478.26
5	230	0.07	0.85	25000	16250	3782.61	22332.61
6	230	0.06	0.91	30000	19500	4043.48	21743.48
7	230	0.03	0.94	35000	22750	4173.91	19823.91
8	230	0.03	0.98	40000	26000	4326.09	18126.09
9	230	0.02	1.00	45000	29250	4413.04	15763.04
10	230	0.00	1.00	50000	32500	4434.78	12734.78

```
knitr::kable(table.RFM, digits=2, caption = "RFM model")
```

Table 15: RFM model

decil	count	mean.choice	mean.RFM	acumul	mailed	cost	sold	profit
1	230	0.03	86.65	0.03	5000	3250	152.17	-1697.83
2	230	0.10	82.22	0.14	10000	6500	608.70	-291.30
3	230	0.06	76.30	0.20	15000	9750	891.30	-658.70
4	230	0.00	75.00	0.20	20000	13000	891.30	-3908.70
5	230	0.21	72.07	0.41	25000	16250	1826.09	2376.09
6	230	0.15	66.96	0.56	30000	19500	2478.26	5778.26
7	230	0.00	65.00	0.56	35000	22750	2478.26	2528.26
8	230	0.18	63.83	0.74	40000	26000	3260.87	7260.87
9	230	0.10	57.22	0.83	45000	29250	3695.65	8445.65
10	230	0.17	52.00	1.00	50000	32500	4434.78	12734.78