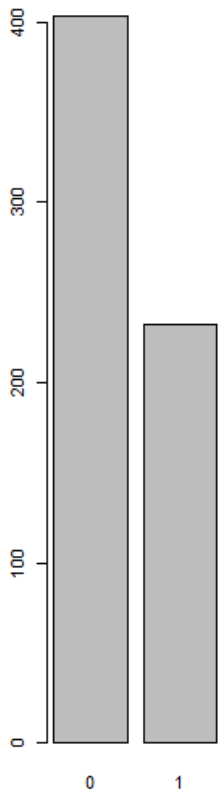
```
> library(dplyr)
>
> setwd('C:/Users/Lenovo/Downloads')
>
> data=read.csv('Forddata.csv', header=T)
> head(data)
                  Date.bought.Customer.Bought.Amount
15.08.1990;113;other;27                          29
16.08.1990;64;other;24                           48
18.08.1990;49;ford;37                            21
18.08.1990;54;ford;40                            95
18.08.1990;83;other;9                             2
20.08.1990;42;other;12                           67
>
> #I realized header not seem ok so before I change it I try with declaring separater and it worked.
>
> data=read.csv('Forddata.csv', sep=';')
> head(data)
  Date.bought Customer Bought Amount
1  15.08.1990      113  other  27,29
2  16.08.1990       64  other  24,48
3  18.08.1990       49   ford  37,21
4  18.08.1990       54   ford  40,95
5  18.08.1990       83  other   9,02
6  20.08.1990       42  other  12,67
>
> #lets fix date s format
> data$Date.bought <- as.Date(as.character(data$Date.bought),"%d.%m.%Y")
>
> data$Amount <- as.double(data$Amount, ",")
> head(data)
  Date.bought Customer Bought Amount
1  1990-08-15      113  other    292
2  1990-08-16       64  other    254
3  1990-08-18       49   ford    436
4  1990-08-18       54   ford    469
5  1990-08-18       83  other    571
6  1990-08-20       42  other     49
>
> #lets see ford and others as bar graph
>
> boughtdata <- table(data$Bought)
> barplot(boughtdata)
```




```
> #we can clearly see that customers choose other ones almost twice as ford. I'll change ford as 1 other as 0
> data$Bought <- ifelse(data$Bought == "ford",1,0)
> head(data)
  Date.bought Customer Bought Amount
1  1990-08-15      113      0    292
2  1990-08-16       64      0    254
3  1990-08-18       49      1    436
4  1990-08-18       54      1    469
5  1990-08-18       83      0    571
6  1990-08-20       42      0     49
>
> data$Date.bought <- as.Date(data$Date.bought)
> class(data$Date.bought)
[1] "Date"
>
> data$last_date <- as.Date("20-10-2012")
>
> #find the last date of data
>
> data <- data %>% mutate("last_date" = max(Date.bought))
>
> data$recency <- difftime(data$last_date, data$Date.bought, units = "days")
> #as we can see our last date is in 2012
> tail(data)
    Date.bought Customer Bought Amount  last_date  recency
630  2010-06-07       54      0    281 2012-04-05 668 days
631  2010-07-19      100      0     69 2012-04-05 626 days
632  2010-09-29      104      0     41 2012-04-05 554 days
```

```
633  2011-01-22       13     0    350 2012-04-05 439 days
634  2011-06-07       16     1    505 2012-04-05 303 days
635  2012-04-05        2     1    225 2012-04-05   0 days
>
> #lets see our customer details
> data <- data %>% arrange(Customer)
> head(data)
  Date.bought Customer Bought Amount  last_date   recency
1  1990-09-24        1      1    538 2012-04-05 7864 days
2  1993-04-27        1      1    570 2012-04-05 6918 days
3  1995-06-01        1      1    506 2012-04-05 6153 days
4  1999-05-17        1      1    315 2012-04-05 4707 days
5  2002-05-11        1      0    578 2012-04-05 3617 days
6  2004-07-23        1      0    246 2012-04-05 2813 days
>
> data <- data %>% group_by(Customer) %>% mutate("frequency" = length(Bought))
>
> data <- data %>% group_by(Customer) %>% mutate("monetary" = sum(Amount))
>
> #data <- data[!(duplicated(data$Customer)),] will be only one customer but date will be according to first purc
hase
>
> frequency <- as.data.frame(table(data[,2]))
>
> #lets create our rfm
>
> data$r <- ntile(data$recency,4)
> data$f <- ntile(data$frequency,4)
> data$m<- ntile(data$monetary,4)
>
> head(data)
# A tibble: 6 x 11
# Groups:   Customer [1]
  Date.bought Customer Bought Amount last_date  recency   frequency monetary     r     f     m
  <date>       <int>  <dbl>  <dbl> <date>     <drtn>       <int>   <dbl> <int> <int> <int>
1 1990-09-24        1      1    538 2012-04-05 7864 days        7    2768     4     3     4
2 1993-04-27        1      1    570 2012-04-05 6918 days        7    2768     4     3     4
3 1995-06-01        1      1    506 2012-04-05 6153 days        7    2768     3     3     4
4 1999-05-17        1      1    315 2012-04-05 4707 days        7    2768     2     3     4
5 2002-05-11        1      0    578 2012-04-05 3617 days        7    2768     1     3     4
6 2004-07-23        1      0    246 2012-04-05 2813 days        7    2768     1     3     4
> #I'm going to play with segments since I want best user to be 111
> #for f
>
> data$f <- with(data, ifelse(f == 1,4, ifelse(f == 2,3, ifelse(f == 3,2, ifelse(f == 4,1, "other")))))
> data$f <- as.integer(data$f)
> #for m
>
> data$m <- with(data, ifelse(m == 1,4, ifelse(m == 2,3, ifelse(m == 3,2, ifelse(m == 4,1, "other")))))
> data$m <- as.integer(data$m)
>
> #we can merge them for betterlook and calculate our total score also
>
> data$rfm_class <- with(data, paste(r,f,m,sep = ""))
> head(data)
# A tibble: 6 x 12
# Groups:   Customer [1]
```

```
  Date.bought Customer Bought Amount last_date  recency   frequency monetary     r     f     m rfm_class
  <date>       <int>  <dbl>  <dbl> <date>     <drtn>       <int>   <dbl> <int> <int> <int> <chr>
1 1990-09-24       1     1    538 2012-04-05 7864 days       7    2768     4     2     1 421
2 1993-04-27       1     1    570 2012-04-05 6918 days       7    2768     4     2     1 421
3 1995-06-01       1     1    506 2012-04-05 6153 days       7    2768     3     2     1 321
4 1999-05-17       1     1    315 2012-04-05 4707 days       7    2768     2     2     1 221
5 2002-05-11       1     0    578 2012-04-05 3617 days       7    2768     1     2     1 121
6 2004-07-23       1     0    246 2012-04-05 2813 days       7    2768     1     2     1 121
>
> tail(data)
# A tibble: 6 x 12
# Groups:   Customer [2]
  Date.bought Customer Bought Amount last_date  recency   frequency monetary     r     f     m rfm_class
  <date>       <int>  <dbl>  <dbl> <date>     <drtn>       <int>   <dbl> <int> <int> <int> <chr>
1 2001-12-24     116     0    258 2012-04-05 3755 days       6    1714     1     2     2 122
2 2004-03-25     116     0     87 2012-04-05 2933 days       6    1714     1     2     2 122
3 1990-11-05     117     0     40 2012-04-05 7822 days       4     562     4     4     4 444
4 1993-09-30     117     0    111 2012-04-05 6762 days       4     562     3     4     4 344
5 1997-10-11     117     0    313 2012-04-05 5290 days       4     562     2     4     4 244
6 2000-05-20     117     0     98 2012-04-05 4338 days       4     562     2     4     4 244
> data$Total_Score <- c(100*data$recency + 10*data$frequency+data$monetary)
>
>
> #forgot to change rfm_class and recency to numeric
>
> data$recency<- as.numeric(data$recency)
> data$rfm_class<- as.numeric(data$rfm_class)
> data$Total_Score<-as.numeric(data$Total_Score)
> #now we can see our results
> par(mfrow = c(1,4))
> hist(data$recency)
> hist(data$frequency)
> hist(data$monetary)
> hist(data$Total_Score)
>
> #as we design, I want to target customers who is recency is high first I try with rfm
> target <- data[data$rfm_class>=441,]
> dim(target)
[1] 53 13
> head(target)
# A tibble: 6 x 13
# Groups:   Customer [4]
  Date.bought Customer Bought Amount last_date  recency frequency monetary     r     f     m rfm_class Total_Sco
re
  <date>       <int>  <dbl>  <dbl> <date>     <dbl>     <int>   <dbl> <int> <int> <int>     <dbl>      <dbl>
1 1990-11-16       3     0    309 2012-04-05    7811         4    1267     4     4     4       444     782407
2 1990-09-07       5     0    566 2012-04-05    7881         4    1080     4     4     4       444     789220
3 1993-01-07       5     0     72 2012-04-05    7028         4    1080     4     4     4       444     703920
4 1990-08-28       7     0    204 2012-04-05    7891         4     828     4     4     4       444     789968
5 1990-10-20       9     1    482 2012-04-05    7838         4    2100     4     4     1       441     785940
6 1993-03-21       9     1    565 2012-04-05    6955         4    2100     4     4     1       441     697640
> #then score
> target <- data[data$Total_Score>=700000,]
> dim(target)
[1] 147  13
> head(target)
# A tibble: 6 x 13
```

Groups: Customer [5]

	Date.bought	Customer	Bought	Amount	last_date	recency	frequency	monetary	r	f	m	rfm_class	Total_Score
	<date>	<int>	<dbl>	<dbl>	<date>	<dbl>	<int>	<dbl>	<int>	<int>	<int>	<dbl>	<dbl>
1	1990-09-24	1	1	538	2012-04-05	7864	7	2768	4	2	1	421	789238
2	1990-10-03	2	1	44	2012-04-05	7855	7	2332	4	2	1	421	787902
3	1990-11-16	3	0	309	2012-04-05	7811	4	1267	4	4	4	444	782407
4	1990-08-23	4	1	238	2012-04-05	7896	7	2044	4	2	2	422	791714
5	1992-12-02	4	1	97	2012-04-05	7064	7	2044	4	2	2	422	708514
6	1990-09-07	5	0	566	2012-04-05	7881	4	1080	4	4	4	444	789220