startDate <- "2017-01-01"

endDate <- "2017-07-20"

InvalidBuyers <- c("extra", "Recharge", "cash", "lkolp")

Recency – 3 level based on equal percentile

Frequency – 3 levels : 1 st is with frequency < 3, rest two are 50% of rest of the data

Monetary – 3 levels based on cumulative amount (30%, 50%, 20%)

> summary(Invoice)

InvoiceID UID InvoiceDate ClientID GrandTotal

Min. : 634810 Min. :25457 Length:44684 Min. : 0 Min. :-1606.0

1st Qu.: 645981 1st Qu.:25457 Class :character 1st Qu.:1104957 1st Qu.: 46.0

Median :2387482 Median :25457 Mode :character Median :2140470 Median : 113.0

Mean :2897018 Mean :25457 Mean :1848500 Mean : 261.6

3rd Qu.:4573610 3rd Qu.:25457 3rd Qu.:2140470 3rd Qu.: 289.0

Max. :8267437 Max. :25457 Max. :6240956 Max. :25753.0

NA's :166

StoreId IsCanceled PaymentType CustomerName CustomerMobileNo

5371:44684 FALSE:44671 1 : 294 Length:44684 Min. :1.010e+02

TRUE : 13 2 : 2394 Class :character 1st Qu.:1.230e+02

3 :41995 Mode :character Median :1.237e+08

NA's: 1 Mean :1.601e+09

3rd Qu.:1.237e+09

Max. :9.793e+10

NA's :170

> sapply(Invoice, function(x) sum(is.na(x)))

InvoiceID UID InvoiceDate ClientID GrandTotal

0 0 0 166 0

StoreId IsCanceled PaymentType CustomerName CustomerMobileNo

0 0 1 176 170

[1] "Data taken between 2017-01-01 and 2017-07-20"

[1] "Least Date in Dataset is 2017-02-22 and Max date is 2018-03-26"

[1] "Invoices between 2017-01-01 and 2017-07-20 are 14153"

[1] "\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*"

[1] "Unique Identification : 1347"

[1] "Valid Identification : 1326"

[1] "InValid Identification: 21"

[1] "\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*"

[1] "\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*"

[1] "InValid Invoices Based on Customers: 10166"

[1] "Valid Invoices Based on Customers : 3987"

[1] "\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*"

[1] "\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*"

[1] "InValid Invoices For Analysis: 3793"

[1] "Valid Invoices : 10360"

[1] "\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*"

> InvoiceDE <- CreateDecile(InvoiceDE,"GrandTotal",0.1)

10% 20% 30% 40% 50% 60% 70% 80% 90%

24.0 42.0 66.6 95.0 131.0 187.0 270.0 407.0 704.8

[1] "Dec-GrandTotal"

[1] "Decile 1 : Between -549 and 24 with 1387 Records"

[1] "Decile 2 : Between 24 and 42 with 1439 Records"

[1] "Decile 3 : Between 42 and 66.6000000000004 with 1420 Records"

[1] "Decile 4 : Between 66.6000000000004 and 95 with 1393 Records"

[1] "Decile 5 : Between 95 and 131 with 1429 Records"

[1] "Decile 6 : Between 131 and 187 with 1422 Records"

[1] "Decile 7 : Between 187 and 270 with 1388 Records"

[1] "Decile 8 : Between 270 and 407 with 1438 Records"

[1] "Decile 9 : Between 407 and 704.800000000001 with 1421 Records"

[1] "Decile 10 : Between 704.800000000001 and 25753 with 1416 Records"

> InvoiceDE <- CreateDecile(InvoiceDE,"GrandTotal",0.25)

25% 50% 75%

54 131 325

[1] "Qart-GrandTotal"

[1] "Quartile 1 : Between -549 and 54 with 3519 Records"

[1] "Quartile 2 : Between 54 and 131 with 3549 Records"

[1] "Quartile 3 : Between 131 and 325 with 3537 Records"

[1] "Quartile 4 : Between 325 and 25753 with 3548 Records"

> describe(InvoiceDE$GrandTotal)

vars n mean sd median trimmed mad min max range skew kurtosis se

X1 1 14153 295.44 532.32 131 191.06 142.33 -549 25753 26302 11.69 399.5 4.47

> InvoiceDEValidData <- CreateDecile(InvoiceDE[InvoiceDE$IsValidData==1,],"GrandTotal",0.1)

10% 20% 30% 40% 50% 60% 70% 80% 90%

30 60 90 137 203 299 423 634 1090

[1] "Dec-GrandTotal"

[1] "Decile 1 : Between -500 and 30 with 368 Records"

[1] "Decile 2 : Between 30 and 60 with 385 Records"

[1] "Decile 3 : Between 60 and 90 with 353 Records"

[1] "Decile 4 : Between 90 and 137 with 405 Records"

[1] "Decile 5 : Between 137 and 203 with 385 Records"

[1] "Decile 6 : Between 203 and 299 with 376 Records"

[1] "Decile 7 : Between 299 and 423 with 382 Records"

[1] "Decile 8 : Between 423 and 634 with 378 Records"

[1] "Decile 9 : Between 634 and 1090 with 379 Records"

[1] "Decile 10 : Between 1090 and 25753 with 382 Records"

> InvoiceDEValidData <- CreateDecile(InvoiceDE[InvoiceDE$IsValidData==1,],"GrandTotal",0.25)

25% 50% 75%

75 203 513

[1] "Qart-GrandTotal"

[1] "Quartile 1 : Between -500 and 75 with 921 Records"

[1] "Quartile 2 : Between 75 and 203 with 975 Records"

[1] "Quartile 3 : Between 203 and 513 with 948 Records"

[1] "Quartile 4 : Between 513 and 25753 with 949 Records"

> RFMData <- PrepareRFM(InvoiceDE, endDate,,,,)

[1] "\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*"

[1] "Two Datasets will be created for valid date"

[1] "NameAndPhone : Recency, Frequency and Monetary for the combination of Name and Customet Identification"

[1] "InvoiceRFM : Recency, Frequency and Monetary Customet Identification"

[1] "\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*"

[1] "\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*"

[1] "Total Number of RFM Customers: 1323"

[1] "\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*"

> InvoiceRFM <- CreateDecile(InvoiceRFM,"Recency",0.1)

10% 20% 30% 40% 50% 60% 70% 80% 90%

2.0 18.0 50.6 81.0 97.0 107.0 116.0 125.0 137.0

[1] "Dec-Recency"

[1] "Decile 1 : Between 0 and 2 with 107 Records"

[1] "Decile 2 : Between 2 and 18 with 156 Records"

[1] "Decile 3 : Between 18 and 50.6000000000001 with 134 Records"

[1] "Decile 4 : Between 50.6000000000001 and 81 with 126 Records"

[1] "Decile 5 : Between 81 and 97 with 135 Records"

[1] "Decile 6 : Between 97 and 107 with 120 Records"

[1] "Decile 7 : Between 107 and 116 with 140 Records"

[1] "Decile 8 : Between 116 and 125 with 137 Records"

[1] "Decile 9 : Between 125 and 137 with 120 Records"

[1] "Decile 10 : Between 137 and 148 with 148 Records"

> InvoiceRFM <- CreateDecile(InvoiceRFM,"Recency",0.25)

25% 50% 75%

25.5 97.0 120.0

[1] "Qart-Recency"

[1] "Quartile 1 : Between 0 and 25.5 with 331 Records"

[1] "Quartile 2 : Between 25.5 and 97 with 327 Records"

[1] "Quartile 3 : Between 97 and 120 with 323 Records"

[1] "Quartile 4 : Between 120 and 148 with 342 Records"

> InvoiceRFM <- CreateDecile(InvoiceRFM,"Frequency",0.1)

10% 20% 30% 40% 50% 60% 70% 80% 90%

1 1 1 1 1 2 2 3 5

[1] "Dec-Frequency"

[1] "Decile 1 : Between 1 and 1 with 0 Records"

[1] "Decile 2 : Between 1 and 1 with 0 Records"

[1] "Decile 3 : Between 1 and 1 with 0 Records"

[1] "Decile 4 : Between 1 and 1 with 0 Records"

[1] "Decile 5 : Between 1 and 1 with 0 Records"

[1] "Decile 6 : Between 1 and 2 with 784 Records"

[1] "Decile 7 : Between 2 and 2 with 0 Records"

[1] "Decile 8 : Between 2 and 3 with 242 Records"

[1] "Decile 9 : Between 3 and 5 with 158 Records"

[1] "Decile 10 : Between 5 and 193 with 139 Records"

> InvoiceRFM <- CreateDecile(InvoiceRFM,"Frequency",0.25)

25% 50% 75%

1 1 2

[1] "Qart-Frequency"

[1] "Quartile 1 : Between 1 and 1 with 0 Records"

[1] "Quartile 2 : Between 1 and 1 with 0 Records"

[1] "Quartile 3 : Between 1 and 2 with 784 Records"

[1] "Quartile 4 : Between 2 and 193 with 539 Records"

> InvoiceRFM <- CreateDecile(InvoiceRFM,"MonetaryAvg",0.1)

10% 20% 30% 40% 50% 60% 70% 80% 90%

75.0000 115.1600 175.1714 236.5333 314.6667 417.0667 550.4800 769.2000 1176.1333

[1] "Dec-MonetaryAvg"

[1] "Decile 1 : Between 1 and 75 with 131 Records"

[1] "Decile 2 : Between 75 and 115.16 with 134 Records"

[1] "Decile 3 : Between 115.16 and 175.171428571429 with 132 Records"

[1] "Decile 4 : Between 175.171428571429 and 236.533333333333 with 132 Records"

[1] "Decile 5 : Between 236.533333333333 and 314.666666666667 with 132 Records"

[1] "Decile 6 : Between 314.666666666667 and 417.066666666667 with 133 Records"

[1] "Decile 7 : Between 417.066666666667 and 550.48 with 132 Records"

[1] "Decile 8 : Between 550.48 and 769.2 with 132 Records"

[1] "Decile 9 : Between 769.2 and 1176.13333333333 with 132 Records"

[1] "Decile 10 : Between 1176.13333333333 and 5950 with 133 Records"

> InvoiceRFM <- CreateDecile(InvoiceRFM,"MonetaryAvg",0.25)

25% 50% 75%

149.3333 314.6667 655.0000

[1] "Qart-MonetaryAvg"

[1] "Quartile 1 : Between 1 and 149.333333333333 with 331 Records"

[1] "Quartile 2 : Between 149.333333333333 and 314.666666666667 with 330 Records"

[1] "Quartile 3 : Between 314.666666666667 and 655 with 330 Records"

[1] "Quartile 4 : Between 655 and 5950 with 332 Records"

> InvoiceRFM <- CreateDecile(InvoiceRFM,"MonetaryTotal",0.1)

10% 20% 30% 40% 50% 60% 70% 80% 90%

81.2 152.4 236.0 354.0 520.0 715.4 1022.4 1643.6 2989.8

[1] "Dec-MonetaryTotal"

[1] "Decile 1 : Between 1 and 81.2 with 133 Records"

[1] "Decile 2 : Between 81.2 and 152.4 with 132 Records"

[1] "Decile 3 : Between 152.4 and 236 with 131 Records"

[1] "Decile 4 : Between 236 and 354 with 132 Records"

[1] "Decile 5 : Between 354 and 520 with 133 Records"

[1] "Decile 6 : Between 520 and 715.4 with 133 Records"

[1] "Decile 7 : Between 715.4 and 1022.4 with 132 Records"

[1] "Decile 8 : Between 1022.4 and 1643.6 with 132 Records"

[1] "Decile 9 : Between 1643.6 and 2989.8 with 132 Records"

[1] "Decile 10 : Between 2989.8 and 32932 with 133 Records"

> InvoiceRFM <- CreateDecile(InvoiceRFM,"MonetaryTotal",0.25)

25% 50% 75%

190.5 520.0 1234.0

[1] "Qart-MonetaryTotal"

[1] "Quartile 1 : Between 1 and 190.5 with 331 Records"

[1] "Quartile 2 : Between 190.5 and 520 with 330 Records"

[1] "Quartile 3 : Between 520 and 1234 with 331 Records"

[1] "Quartile 4 : Between 1234 and 32932 with 331 Records"

> ## To analyse cumulative distribution of amount

> InvoiceRFM <- GetCumScore(InvoiceRFM, , 10,NULL)

[1] 6

[1] "Record for first 10 % : 6 with Minimum value as 19997 and maximum value as 32932"

[1] 16

[1] "Record for first 20 % : 16 with Minimum value as 12381 and maximum value as 32932"

[1] 31

[1] "Record for first 30 % : 31 with Minimum value as 7670 and maximum value as 32932"

[1] 50

[1] "Record for first 40 % : 50 with Minimum value as 5980 and maximum value as 32932"

[1] 71

[1] "Record for first 50 % : 71 with Minimum value as 4736 and maximum value as 32932"

[1] 94

[1] "Record for first 60 % : 94 with Minimum value as 3965 and maximum value as 32932"

[1] 119

[1] "Record for first 70 % : 119 with Minimum value as 3325 and maximum value as 32932"

[1] 146

[1] "Record for first 80 % : 146 with Minimum value as 2778 and maximum value as 32932"

[1] 175

[1] "Record for first 90 % : 175 with Minimum value as 2445 and maximum value as 32932"

[1] 1323

[1] "Record for first 100 % : 1323 with Minimum value as 1 and maximum value as 32932"

> InvoiceRFMScore <- GetRScore(InvoiceRFM[,c(1,2,3,4,5,6,7,8,9)],rLowest,rHighest,r,rvector)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.00 25.50 97.00 80.34 120.00 148.00

[1] 67 113

[1] 67 113

[1] "Score-Recency"

[1] "Segment 1 : Between 0 and 67 with 438 Records"

[1] "Segment 2 : Between 67 and 113 with 435 Records"

[1] "Segment 3 : Between 113 and 148 with 450 Records"

> InvoiceRFMScore <- GetFMScore(InvoiceRFMScore,"Frequency",fLowest,fHighest,f,fvector)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 1.000 1.000 2.848 2.000 193.000

[1] 3 6

[1] 3 6

[1] "Score-Frequency"

[1] "Segment 1 : Between 1 and 3 with 1026 Records"

[1] "Segment 2 : Between 3 and 6 with 198 Records"

[1] "Segment 3 : Between 6 and 193 with 99 Records"

> InvoiceRFMScore <- GetCumScore(InvoiceRFMScore, , ,c(25,35,40))

[1] 28

[1] "Record for first 25 % : 28 with Minimum value as 8299 and maximum value as 32932"

[1] 118

[1] "Record for first 60 % : 118 with Minimum value as 3358 and maximum value as 32932"

[1] 1323

[1] "Record for first 100 % : 1323 with Minimum value as 1 and maximum value as 32932"