Ex6_Team3

Naga Soundari Balamurugan

November 13, 2018

```
#Required libraries
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(gmodels)

source("RFM_Functions.R")
```

```
tuscan_data <- read.csv("Tuscan_RFM_Data_R.csv")
liftGainData <- read.csv("Lifts_and_Gain_Data_R.csv")

colnames(liftGainData) <- c("RFM", "Customers", "Buyers")
liftGainData = liftGainData[,1:3]

liftGainData$ResponseRate <- liftGainData$Buyers/liftGainData$Customers
```

Part I: Preliminary and Quintile Analysis (Independent RFM)

1. What percent of customers responded (i.e. bought anything) from this catalog?

```
total_no_customers <- nrow(tuscan_data)
responded_customers <- tuscan_data %>% filter(buyer == 1)
no_res_customers <- nrow(responded_customers)

percent_responded <- (no_res_customers/total_no_customers) * 100
percent_responded
```

```
## [1] 2.455697
```

The percent of customer who responded from the catalog is 2.4556%

2. Of those who bought, what was the average dollars ordered from this catalog?

```
avg_dollar <- mean(responded_customers$dollars)
avg_dollar
```

```
## [1] 104.2429
```

Average dollars ordered from the catalog is $104.2429.

3. Create quintile (i.e., split by 5) RFM scores for recency, frequency and monetary.

```
data_rfm <- data.frame(ID = tuscan_data$ID, Recency = tuscan_data$last,
                       Frequency = tuscan_data$numords, Monetary = tuscan_data$totdol)
RFM_score <- getIndependentScore(data_rfm)
```

4. Create a bar chart showing the response rate (i.e., the proportion of customers who bought something) by recency quintile.

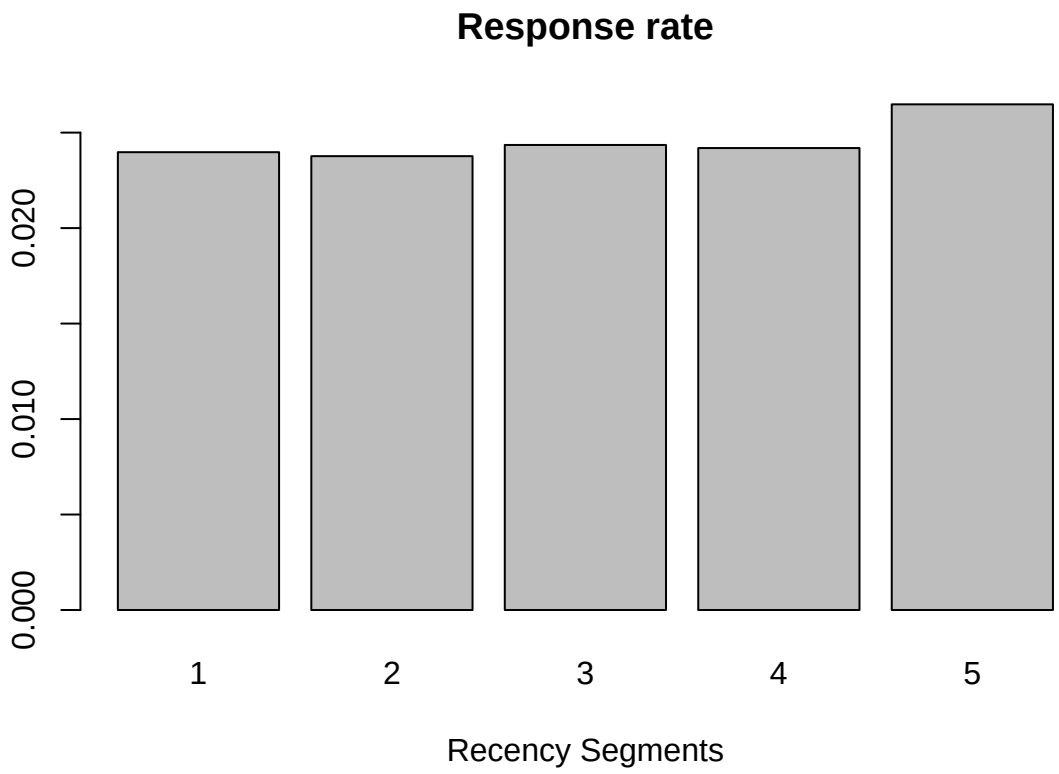
```
# Need to sort RFM.score by Customer ID
# Note that tuscan_data is sorted by ID!
sortedRFM <- RFM_score[order(RFM_score$ID),]


# Crosstab of Recency Score vs. Buyer (did or did not buy offer)
data_crosstab_Recency <- CrossTable(sortedRFM$R_Score, tuscan_data$buyer, prop.r = TRUE,
      prop.c = FALSE, prop.t = FALSE, prop.chisq = FALSE, dnn = c("Recency", "Response"))
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |-------------------------|
##
##
## Total Observations in Table:  96551
##
##
##              | Response
##      Recency |         0 |         1 | Row Total |
## -------------|-----------|-----------|-----------|
##            1 |     18727 |       460 |     19187 |
##              |     0.976 |     0.024 |     0.199 |
## -------------|-----------|-----------|-----------|
##            2 |     18854 |       459 |     19313 |
##              |     0.976 |     0.024 |     0.200 |
## -------------|-----------|-----------|-----------|
##            3 |     18909 |       472 |     19381 |
##              |     0.976 |     0.024 |     0.201 |
## -------------|-----------|-----------|-----------|
##            4 |     18757 |       465 |     19222 |
##              |     0.976 |     0.024 |     0.199 |
## -------------|-----------|-----------|-----------|
```

```
##             5 |     18933 |       515 |     19448 |
##               |     0.974 |     0.026 |     0.201 |
## -------------|-----------|-----------|-----------|
## Column Total |     94180 |      2371 |     96551 |
## -------------|-----------|-----------|-----------|
##
##
```

```r
# Bar Plot
barplot(data_crosstab_Recency$prop.row[,2], main="Response rate",
    xlab="Recency Segments")
```



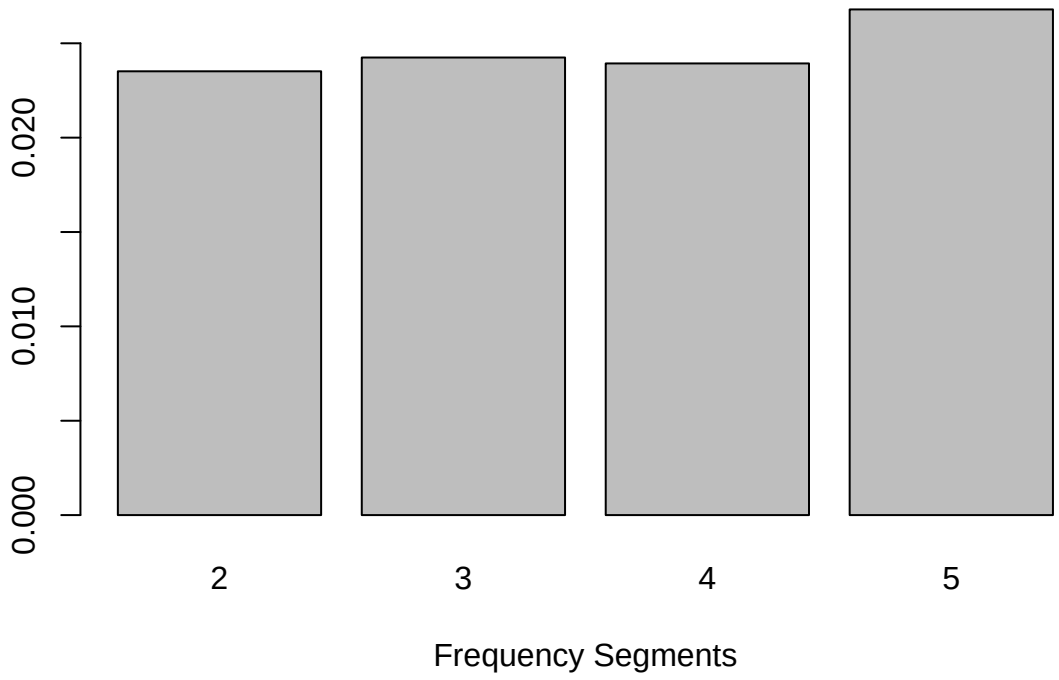
5. Create a bar chart showing the response rate by frequency quintile.

```r
# Crosstab of Recency Score vs. Buyer (did or did not buy offer)
data_crosstab_Frequency <- CrossTable(sortedRFM$F_Score, tuscan_data$buyer, prop.r = TRUE,
      prop.c = FALSE, prop.t = FALSE, prop.chisq = FALSE, dnn = c("Frequency", "Response"))
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |-------------------------|
```

```
##
##
## Total Observations in Table:  96551
##
##
##              | Response
##    Frequency |         0 |         1 | Row Total |
## -------------|-----------|-----------|-----------|
##            2 |     34093 |       821 |     34914 |
##              |     0.976 |     0.024 |     0.362 |
## -------------|-----------|-----------|-----------|
##            3 |     18351 |       456 |     18807 |
##              |     0.976 |     0.024 |     0.195 |
## -------------|-----------|-----------|-----------|
##            4 |     18272 |       448 |     18720 |
##              |     0.976 |     0.024 |     0.194 |
## -------------|-----------|-----------|-----------|
##            5 |     23464 |       646 |     24110 |
##              |     0.973 |     0.027 |     0.250 |
## -------------|-----------|-----------|-----------|
## Column Total |     94180 |      2371 |     96551 |
## -------------|-----------|-----------|-----------|
##
##
```

```r
# Bar Plot
barplot(data_crosstab_Frequency$prop.row[,2], main="Response rate",
    xlab="Frequency Segments")
```



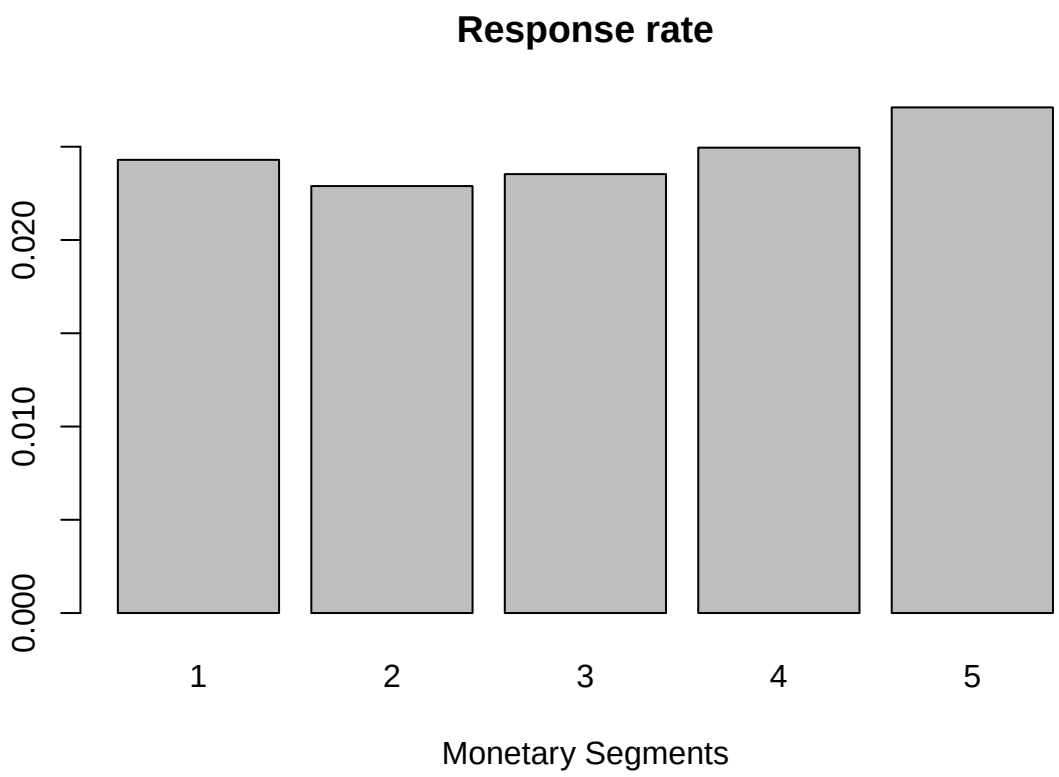
6. Create a bar chart showing the response rate by monetary quintile.

```
# Crosstab of Recency Score vs. Buyer (did or did not buy offer)
data_crosstab_Monetary <- CrossTable(sortedRFM$M_Score, tuscan_data$buyer, prop.r = TRUE,
      prop.c = FALSE, prop.t = FALSE, prop.chisq = FALSE, dnn = c("Monetary", "Response"))
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |-------------------------|
##
##
## Total Observations in Table:  96551
##
##
##              | Response
##     Monetary |         0 |         1 | Row Total |
## -------------|-----------|-----------|-----------|
##            1 |     18754 |       467 |     19221 |
##              |     0.976 |     0.024 |     0.199 |
## -------------|-----------|-----------|-----------|
##            2 |     18909 |       443 |     19352 |
```

```
##              |     0.977 |     0.023 |     0.200 |
## -------------|-----------|-----------|-----------|
##            3 |     18839 |       454 |     19293 |
##              |     0.976 |     0.024 |     0.200 |
## -------------|-----------|-----------|-----------|
##            4 |     18837 |       482 |     19319 |
##              |     0.975 |     0.025 |     0.200 |
## -------------|-----------|-----------|-----------|
##            5 |     18841 |       525 |     19366 |
##              |     0.973 |     0.027 |     0.201 |
## -------------|-----------|-----------|-----------|
## Column Total |     94180 |      2371 |     96551 |
## -------------|-----------|-----------|-----------|
##
##
```

```r
# Bar Plot
barplot(data_crosstab_Monetary$prop.row[,2], main="Response rate",
    xlab="Monetary Segments")
```



7. What do the above bar charts reveal about the likelihood of response and the size of the order across the different recency, frequency, and monetary quintiles?

There is no significant difference between the quintiles. The quintiles based on all three factors have a similar response rate. Among all 3 factors, the segment 5 is highest and segment 2 is lowest

only with a small difference. Thus, the likelihood of response does not get affected much by these factors.

Part II: Profitability Analysis

1. If the firm would mail to all 96,551 customers, please calculate:

(a) the gross profit in dollars,

(b) the gross profit as a % of gross sales, and

(c) the return on marketing expenditures (gross profit/cost to mail catalogs)

```
gross_sales <- sum(responded_customers$dollars)
catalog_cost <- 96551 * 1
COGS <- 0.5 * gross_sales

gross_profit_dollars <- gross_sales - catalog_cost - COGS
gross_profit_dollars
```

```
## [1] 27029
```

```
gross_profit_percent <- (gross_profit_dollars/gross_sales) * 100
gross_profit_percent
```

```
## [1] 10.93583
```

```
ret_expenditure <- gross_profit_dollars/catalog_cost
ret_expenditure
```

```
## [1] 0.2799453
```

Gross profit in dollars - $27029 Gross profit in percentage - 10.93583% Return on marketing expenditures - 0.2799453

2. What is the breakeven response rate?

```
#Breakeven response rate = (Cost of Marketing)/(Selling Price - Wholesale price - expenditure)

profit <- (no_res_customers * avg_dollar) - COGS

breakeven_res_rate <- (catalog_cost/profit*percent_responded)
breakeven_res_rate
```

```
## [1] 1.918595
```

Breakeven response rate - 1.918595%

3. You can use the Data from the Lifts&Gains Excel Sheet provided or you can calculate the segment performance data by yourself (Basically, a cross-tab as discussed in class). Based on this new dataset (Lifts&Gains or your own), please answer the following questions

Determine which RFM segments have response rates exceeding the breakeven rate.

```
# data_crosstab_resRate <- CrossTable(sortedRFM$Total_Score, tuscan_data$buyer, prop.r = TRUE,
#       prop.c = FALSE, prop.t = FALSE, prop.chisq = FALSE, dnn = c("RFM Segment", "Response"))
#
# filteredSegments <- data_crosstab_resRate$prop.row[,2] > 0.0191
#
# filteredSegmentsDF <- as.data.frame(filteredSegments)
# filteredSegmentsDF$Segment <-  rownames(filteredSegmentsDF)
#
# goodResRate <- filteredSegmentsDF %>% filter(filteredSegments == TRUE)

filteredSegments <- liftGainData %>% filter(ResponseRate > 0.0191)
filteredSegments
```

```
##    RFM Customers Buyers ResponseRate
## 1  555      5760    340   0.05902778
## 2  554      1751     79   0.04511708
## 3  553       434     18   0.04147465
## 4  552        69      2   0.02898551
## 5  545       638     16   0.02507837
## 6  544      1553     55   0.03541533
## 7  543      1122     40   0.03565062
## 8  542       552     20   0.03623188
## 9  541        93      2   0.02150538
## 10 534       654     16   0.02446483
## 11 533      1048     31   0.02958015
## 12 532       941     28   0.02975558
## 13 531       422     12   0.02843602
## 14 525        15      1   0.06666667
## 15 524       231      9   0.03896104
## 16 523       692     14   0.02023121
## 17 522      1360     37   0.02720588
## 18 521      1979     55   0.02779181
## 19 455      3877    190   0.04900696
## 20 454      1566     63   0.04022989
## 21 453       416     20   0.04807692
## 22 452        65      2   0.03076923
## 23 445       599     14   0.02337229
## 24 444      1457     40   0.02745367
## 25 443      1182     38   0.03214890
## 26 442       585     16   0.02735043
## 27 441       124      4   0.03225806
## 28 434       700     21   0.03000000
## 29 431       553     16   0.02893309
## 30 425        42      1   0.02380952
## 31 422      1712     40   0.02336449
## 32 355      2501     87   0.03478609
```

```
## 33 354      1164     32   0.02749141
## 34 353       316      9   0.02848101
## 35 345       558     11   0.01971326
## 36 344      1451     39   0.02687802
## 37 341        96      2   0.02083333
## 38 333      1273     26   0.02042419
## 39 332      1117     23   0.02059087
## 40 325        43      1   0.02325581
## 41 255      1851     50   0.02701243
## 42 254       927     20   0.02157497
## 43 253       272      6   0.02205882
## 44 245       474     13   0.02742616
## 45 243      1037     22   0.02121504
## 46 241        80      4   0.05000000
## 47 231       610     18   0.02950820
## 48 223      1338     27   0.02017937
## 49 155      1633     39   0.02388242
## 50 154       985     28   0.02842640
## 51 143      1161     32   0.02756245
## 52 133      1465     29   0.01979522
```

Determine the number of customers belonging to these profitable segments.

Determine the number of buyers belonging to these profitable segments.

```
cust_count <- sum(filteredSegments$Customers)
buyer_count <- sum(filteredSegments$Buyers)
cust_count
```

```
## [1] 52544
```

```
buyer_count
```

```
## [1] 1758
```

> No of customers belonging to profitable segments: 52544 No of buyers belonging to profitable segments: 1758

What would the

(a) the gross profit in dollars,

(b) the gross profit as a % of gross sales, and

(c) the return on marketing expenditures (gross profit/cost to mail catalogs) have been as a result of mailing the catalog only to those customers in the RFM cells with response rates exceeding the breakeven?

```
gross_revenue_RFM <- avg_dollar * no_res_customers

gross_profit_RFM_dollars <- gross_revenue_RFM - (0.5 * gross_revenue_RFM) - catalog_cost
gross_profit_RFM_dollars
```

```
## [1] 27029
```

```
gross_profit_RFM_percent <- gross_profit_RFM_dollars/gross_revenue_RFM
gross_profit_RFM_percent
```

```
## [1] 0.1093583
```

```
ret_expenditure_RFM <- gross_profit_RFM_dollars/catalog_cost
ret_expenditure_RFM
```

```
## [1] 0.2799453
```

Gross profit in dollars for selected RFM segments: $27029 Gross profit in percentage for selected RFM segments: 10.93583% The return on marketing expenditures for selected RFM segments: 0.2799453

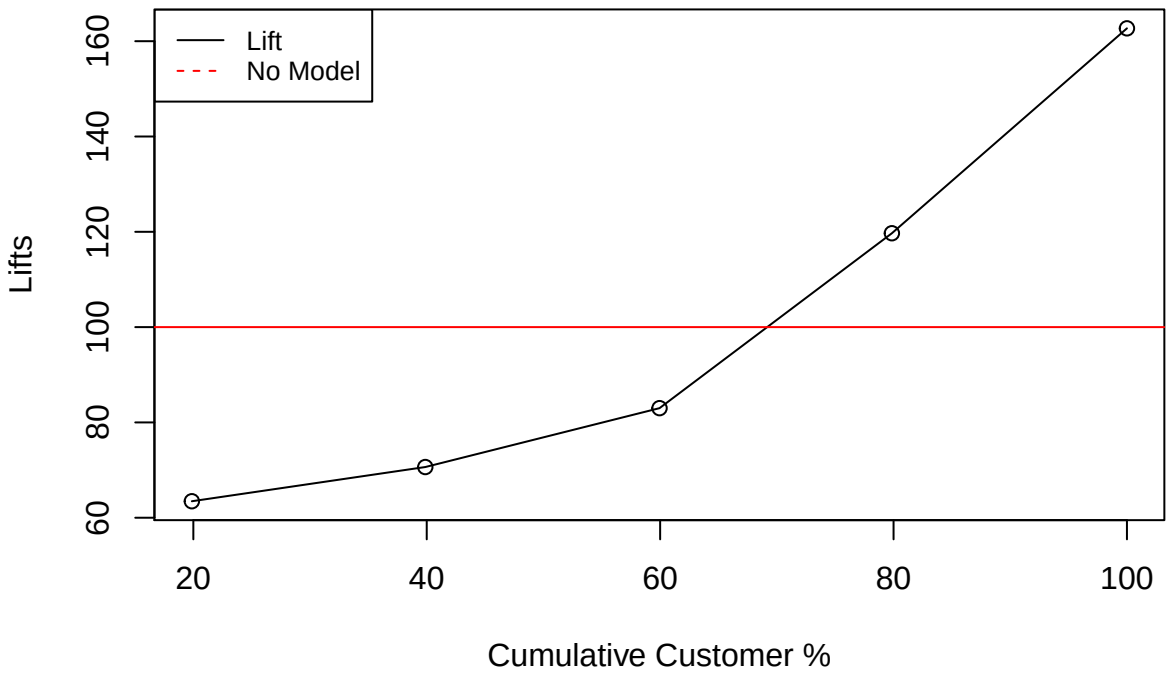
Finally, generate a Lift&Gains chart as discussed in class comparing the RFM model to the no-model scenario.

```
liftGainData$quintile = floor(liftGainData$RFM/100)
pivot_lift = aggregate(liftGainData[,c('Customers', 'Buyers')], by=list(liftGainData$quintile), FUN=sum)
pivot_lift$cuml_cust = cumsum(pivot_lift$Customers)
pivot_lift$cuml_cust_per = pivot_lift$cuml_cust/pivot_lift$cuml_cust[5]*100
pivot_lift$cuml_buy = cumsum(pivot_lift$Buyers)
pivot_lift$cuml_buy_per = pivot_lift$cuml_buy/pivot_lift$cuml_buy[5]*100
pivot_lift$resrate = pivot_lift$Buyers / pivot_lift$Customers*100
total_resrate = sum(pivot_lift$Buyers)/sum(pivot_lift$Customers)*100
pivot_lift$Lift = pivot_lift$resrate/total_resrate*100
pivot_lift$gain_per = pivot_lift$Buyers/sum(pivot_lift$Buyers)*100
pivot_lift$gain_per_cum = cumsum(pivot_lift$gain_per)
pivot_lift
```

```
##   Group.1 Customers Buyers cuml_cust cuml_cust_per cuml_buy cuml_buy_per
## 1       1     19187    299     19187      19.87240      299     12.61071
## 2       2     19313    335     38500      39.87530      634     26.73977
## 3       3     19381    395     57881      59.94863     1029     43.39941
## 4       4     19222    565     77103      79.85728     1594     67.22902
## 5       5     19448    777     96551     100.00000     2371    100.00000
##    resrate      Lift gain_per gain_per_cum
## 1 1.558347  63.45843 12.61071     12.61071
## 2 1.734583  70.63506 14.12906     26.73977
## 3 2.038079  82.99389 16.65964     43.39941
## 4 2.939340 119.69475 23.82961     67.22902
## 5 3.995269 162.69391 32.77098    100.00000
```

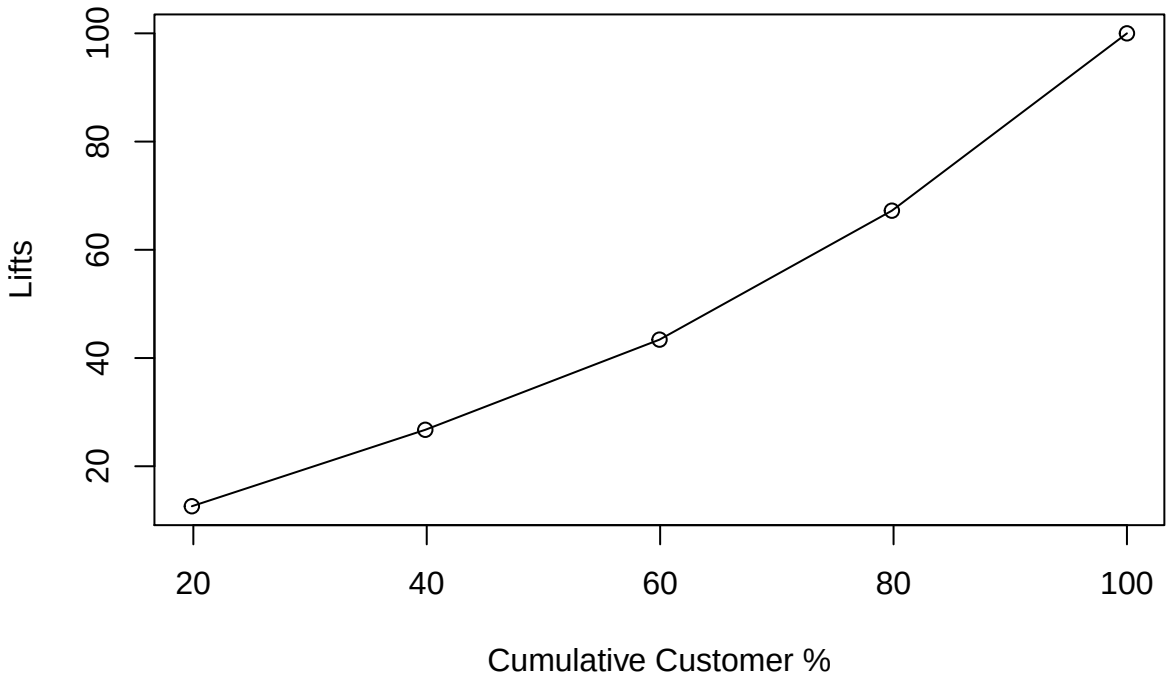
Plotting a banana plot

```
options(warn=-1)
plot(pivot_lift$cuml_cust_per, pivot_lift$Lift, xlab = 'Cumulative Customer %', ylab = 'Lifts')
lines(pivot_lift$cuml_cust_per, pivot_lift$Lift, type = 'l')
abline(h = 100, col = "red")
legend("topleft", legend=c("Lift", "No Model"),
       col=c("black", "red"), lty=1:2, cex=0.8)
```



plot2

```
options(warn=-1)
plot(pivot_lift$gain_per_cum ~ pivot_lift$cuml_cust_per, xlab = 'Cumulative Customer %', ylab = 'Lifts'
lines(pivot_lift$gain_per_cum ~ pivot_lift$cuml_cust_per, type = 'l')
```


Lifts
20
40
60
80
100
20
40
60
80
100
Cumulative Customer %