Homework #4: Customer Valuation

Raveena Kamal

Remember

1. Your task is to fill in all R code blocks that currently contain "#TBD" comments. Similarly, insert text responses wherever you see *TBD* in the markdown file.
2. PLEASE ADD YOUR NAME TO THE AUTHOR LINE ABOVE

Overview & Instructions

For homework, you will:

1) Compute customer lifetime value (CLV) measures for a women's apparel brand by market segment

- The workshop makes use of two data files:
  - `apparel_customer_revenue.csv` – a panel dataset containing observations of total annual revenue for a sample of 1000 customers, over a period of 10 years.
  - `apparel_customer_demogs.csv` – demographic and behavioral data for each of the 1000 customers sampled
- Notes:
  - Sampled customers are from the same "cohort", meaning they all became customers in the same year (0).
  - Years are indexed 0 to 9 for consistency with CLV calculations on existing customers. That is, we consider the "present" (year 0) to be the end of the period (year) in which the customer is acquired, and we calculate lifetime value with respect to this point in time.
  - When calculated in this manner, the CLV represents the 10-year lifetime value of the customer (the present year, plus 9 future years). That is, the CLV is the present value of a 10 year profit stream that begins immediately upon the acquisition of the customer.
  - CLV can also be interpreted as the maximum cost a firm should be willing to pay to acquire a customer, assuming the firm wants to break-even over the horizon of the CLV calculation (10 years, present plus 9 future years). The variables in the `apparel_customer_revenue.csv` are:

Variable	Description
`iid`	Identifier for customer
`revenue_0`	total dollars spent in year 0
`revenue_1`	total dollars spent in year 1
`revenue_2`	total dollars spent in year 2

Variable	Description
revenue_3	total dollars spent in year 3
revenue_4	total dollars spent in year 4
revenue_5	total dollars spent in year 5
revenue_6	total dollars spent in year 6
revenue_7	total dollars spent in year 7
revenue_8	total dollars spent in year 8
revenue_9	total dollars spent in year 9

The variables in the apparel_customer_demogs.csv are:

Variable	Description
iid	Identifier for customer
spend_online0	dollars spent in year 0 on online purchases
spend_retail0	dollars spent in year 0 on retail purchases
age	customer age
male	1 = if consumer is male
white	proportion of households in customer zip code that are white
college	proportion of households in customer zip code that have college
hh_inc	median income of households in customer zip code ('000)
segment	the segment a customer is in according to cluster analysis

Homework task workflow

1. Setup
    1. Load and summarize data files
2. Calculation of CLV by (pre-determined) segments
    1. Simple method
    2. Cohort method

1.1) Download data & R Markdown file

If you have not already done so, download the data files: apparel_customer_revenues.csv and apparel_customer_demogs.csv from Canvas.

1.2) Load and summarize data

First, load the revenue data into a dataframe named DF_rev. Use head() and summary() to visualize the first few rows and to summarize the variables.

```
DF_rev<-read.csv('/Users/raveena/Desktop/Classroom - R/Marketing Analytics/data/apparel_customer_revenue
head(DF_rev)
```

```
##   iid revenue_0 revenue_1 revenue_2 revenue_3 revenue_4 revenue_5 revenue_6
## 1  14    132.98    216.21    169.94     76.23    172.05      0.00      0.00
## 2  19    171.98    153.77     66.62     58.45    228.92    149.24    161.57
## 3  20     92.00    100.30     94.18    100.28     87.62     82.51    135.57
```

```
## 4  27     49.95      0.00      0.00      0.00      0.00      0.00      0.00
## 5  58    367.95    381.21    467.38      0.00      0.00      0.00      0.00
## 6  77     85.97    109.93     70.61     77.57     44.16    123.18      0.00
##   revenue_7 revenue_8 revenue_9
## 1      0.00      0.00      0.00
## 2    179.18    142.36      0.00
## 3    117.32     52.36     95.87
## 4      0.00      0.00      0.00
## 5      0.00      0.00      0.00
## 6      0.00      0.00      0.00
```

```
summary(DF_rev)
```

```
##       iid          revenue_0         revenue_1         revenue_2
##  Min.   :   14   Min.   :   2.47   Min.   :   0.00   Min.   :   0.00
##  1st Qu.: 2946   1st Qu.:  33.99   1st Qu.:  13.26   1st Qu.:   0.00
##  Median : 5430   Median :  64.00   Median :  47.66   Median :  28.50
##  Mean   : 5463   Mean   : 150.44   Mean   : 128.64   Mean   : 103.01
##  3rd Qu.: 8110   3rd Qu.: 146.50   3rd Qu.: 113.75   3rd Qu.:  89.46
##  Max.   :10589   Max.   :3135.92   Max.   :3577.04   Max.   :5456.27
##    revenue_3         revenue_4         revenue_5         revenue_6
##  Min.   :   0.00   Min.   :   0.00   Min.   :   0.00   Min.   :   0.000
##  1st Qu.:   0.00   1st Qu.:   0.00   1st Qu.:   0.00   1st Qu.:   0.000
##  Median :  11.50   Median :   0.00   Median :   0.00   Median :   0.000
##  Mean   :  78.45   Mean   :  59.19   Mean   :  46.09   Mean   :  34.942
##  3rd Qu.:  66.92   3rd Qu.:  47.05   3rd Qu.:  25.24   3rd Qu.:   4.228
##  Max.   :3241.10   Max.   :1822.99   Max.   :2938.26   Max.   :1662.940
##    revenue_7         revenue_8         revenue_9
##  Min.   :   0.00   Min.   :   0.00   Min.   :   0.00
##  1st Qu.:   0.00   1st Qu.:   0.00   1st Qu.:   0.00
##  Median :   0.00   Median :   0.00   Median :   0.00
##  Mean   :  26.98   Mean   :  23.68   Mean   :  19.46
##  3rd Qu.:   0.00   3rd Qu.:   0.00   3rd Qu.:   0.00
##  Max.   :1554.60   Max.   :1325.45   Max.   :1944.20
```

Next, we load the demographic data into a dataframe named `DF_demogs`. Use `head()` and `summary()` to visualize the first few rows and to summarize the variables.

```
DF_demogs <-read.csv('/Users/raveena/Desktop/Classroom - R/Marketing Analytics/data/apparel_customer_de

head(DF_demogs)
```

```
##   X iid spend_online0 spend_retail0 age     white   college male  hh_inc
## 1 1  14        14.975       118.000  29 0.3241053 0.2868369    0  40.322
## 2 2  19       171.975         0.000  55 0.8723629 0.5170230    0  72.500
## 3 3  20         0.000        92.000  47 0.9808348 0.5246835    0  90.582
## 4 4  27        49.950         0.000  40 0.4638865 0.1825573    0  52.621
## 5 5  58       293.950        74.000  49 0.7607200 0.5241683    0  32.278
## 6 6  77         0.000        85.975  38 0.8887451 0.9465517    0 110.000
##   segment
## 1       4
## 2       2
```

```
## 3       1
## 4       2
## 5       4
## 6       1
```

```
summary(DF_demogs)
```

```
##        X               iid         spend_online0     spend_retail0
##  Min.   :   1.0   Min.   :   14   Min.   :   0.00   Min.   :   0.00
##  1st Qu.: 250.8   1st Qu.: 2946   1st Qu.:   0.00   1st Qu.:   0.00
##  Median : 500.5   Median : 5430   Median :  14.97   Median :  27.71
##  Mean   : 500.5   Mean   : 5463   Mean   :  72.44   Mean   :  78.00
##  3rd Qu.: 750.2   3rd Qu.: 8110   3rd Qu.:  70.72   3rd Qu.:  78.00
##  Max.   :1000.0   Max.   :10589   Max.   :1985.75   Max.   :2421.91
##       age            white          college           male
##  Min.   :18.00   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
##  1st Qu.:33.00   1st Qu.:0.7297   1st Qu.:0.3835   1st Qu.:0.000
##  Median :41.00   Median :0.8550   Median :0.5580   Median :0.000
##  Mean   :40.91   Mean   :0.7993   Mean   :0.5437   Mean   :0.091
##  3rd Qu.:49.00   3rd Qu.:0.9422   3rd Qu.:0.7136   3rd Qu.:0.000
##  Max.   :88.00   Max.   :1.0000   Max.   :1.0000   Max.   :1.000
##      hh_inc           segment
##  Min.   :  2.499   Min.   :1.000
##  1st Qu.: 59.356   1st Qu.:1.000
##  Median : 87.364   Median :2.000
##  Mean   : 96.254   Mean   :2.008
##  3rd Qu.:122.602   3rd Qu.:3.000
##  Max.   :250.001   Max.   :4.000
```

2 Calculation of CLV by (pre-determined) segments

Here we allow for heterogeneous CLV values based upon segmentation assignments. The assumption here is that the segmentation scheme and resulting segment assignments have been made in in advance.

2.1 Simple method

Compute CLV measures by segment, using the assignments now given by the variable `DF_demogs$segment`.

To do this, take the following steps:

1) Combine the revenue (`DF_revenue`) and demographics (`DF_demogs`) dataframes by merging on the variable `iid`. Call the resulting data frame `DF_comb`.
    - As an example, to merge dataframes `DF1` and `DF2` using `id`, use: `merge(DF1,DF2,by="id")`
2) Loop over the number of segments in the data (1 to 4, as seen in DF_comb$segment). For each segment:
    a. Restrict the data to rows from `DF_comb` associated with that segment. The `subset()` command may be useful for this task.
    b. Using the subsetted data (only), compute the per-customer per-period profit (M) for the segment, assuming a profit margin of 40%.

c. Using the subsetted data (only), compute the retention rate for the segment.
d. Compute the segment's CLV using your `CLV_simple()` function from this week's workshop. Again assume a discount rate of r=10% and a CLV horizon of T=9 future periods (10 periods total, indexed 0 to 9).
e. Store the CLV value in a list, indexed by segment number

3) Report the CLV's by segment in a table, with related information about the segments. The table should have the following columns:

- segment average for `spend_online0`
- segment average for `spend_retail0`
- segment average for `age`
- segment average for `white`
- segment average for `college`
- segment average for `male`
- segment average for `hh_inc`
- segment size/share
- CLV
- total value = CLV * segment size

Print the table using 2 decimal place accuracy.

```
DF_comb<-merge(DF_rev,DF_demogs,by="iid")


all_segments<-unique(DF_comb$segment)
rmargin <- 0.4
horizon<-10
return_rate<- 0.1


CLV_simple <- function(profit, retention, r, horizon) {
  clv = 0
  for (t in 0:(horizon-1)) {
    clv = clv + retention^t * (profit/(1+r)^t)
  }
  return(clv)
}


#Answer to parts 2.(a,b,c,d,e)

CLV_df <- rep(NA, length(all_segments))
profit<- rep(NA, length(all_segments))
for (i in 1:length(all_segments)) {
  segment_DF <- subset(DF_comb, segment == i)

  profit[i]<-rmargin * mean(segment_DF$revenue_0)

  retention_rate <- sum(segment_DF$revenue_1 > 0)/
  sum(segment_DF$revenue_0 > 0)

  CLV_df[i] <- CLV_simple(profit = profit[i],
           retention = retention_rate,
```

```
          r = return_rate,
          horizon = horizon)
}


##Answers to part 3
final_averages_DF <- data.frame(matrix(nrow=0, ncol=7))
total_value_simple <- rep(NA, length(all_segments))
segment_size <- rep(NA, length(all_segments))

for (i in 1:length(all_segments)) {
  segment_DF<-subset(DF_comb, segment == i)
  column_names <- colnames(segment_DF[, 14:ncol(segment_DF)-1])
  segment_DF_mean <- data.frame(matrix(colMeans(segment_DF[, 14:ncol(segment_DF)-1]),1))
  colnames(segment_DF_mean) <- c(column_names)
  final_averages_DF <- rbind(final_averages_DF,segment_DF_mean)


  segment_size[i]<-nrow(segment_DF)

  total_value_simple[i] <-CLV_df[i]*segment_size[i]
}


Answer_table<-round(data.frame(segment= all_segments,
                               segment_size = segment_size,
                                 final_averages_DF = final_averages_DF,
                                 total_value = total_value_simple,
                                 CLV = CLV_df), digits = 2)

print(Answer_table)
```

```
##   segment segment_size final_averages_DF.spend_online0
## 1       4          428                            0.02
## 2       2          303                          110.37
## 3       1          102                           73.81
## 4       3          167                          188.38
##   final_averages_DF.spend_retail0 final_averages_DF.age final_averages_DF.white
## 1                           88.89                41.56                    0.82
## 2                            0.03                40.37                    0.79
## 3                           71.72                40.84                    0.71
## 4                          195.40                40.26                    0.82
##   final_averages_DF.college final_averages_DF.male final_averages_DF.hh_inc
## 1                      0.58                   0.00                   102.48
## 2                      0.48                   0.00                    85.08
## 3                      0.50                   0.89                    98.49
## 4                      0.58                   0.00                    99.22
##   total_value    CLV
## 1    58212.95 136.01
## 2    46008.99 151.84
## 3    17168.08 168.31
## 4   106257.78 636.27
```

Discussion:

- How much would you be willing to pay to acquire customers from each of these segments?

The amount to acquire customers from each segment will be equal to their CLV for respective segment. (CLV_df)

- Which segments are expected to be most profitable on a per-customer basis? On a total value basis?
  - What do these results imply for targeting customers?

Segment 4 is most profitable followed by segment 3 then segment 2 then segment 1, on a per-customer basis.

On a total value basis segment 4 is most profitable, followed by segment 1, then segment 2 and then segment 3.

These results imply that the customers in segment 4 are well targeted due to high CLV and high total value, compared with other segments. Also, although segment 1 has highest segment size leading to a second high total value means these customers need to be targeted with better marketing tactics. It could also mean that these might be the ones who are. not relevant for the business.

2.2 Cohort method

Now compute CLV measures by segment, using the cohort method. The general flow of the code is similar to the previous section, but where CLV's are computed using your function `CLV_cohort()`.

Report the CLV's by segment in a table, with related information about the segments. The table should have the following columns: + segment average for `spend_online0` + segment average for `spend_retail0` + segment average for `age` + segment average for `white` + segment average for `college` + segment average for `male` + segment average for `hh_inc` + segment size/share + CLV + total value = CLV * segment size

```
CLV_cohort <- function(profits,r) {
  n_customers = nrow(profits)
  n_years = length(profits)

  # compute average profits by year
  avgRev = colMeans(profits)


  # compute CLV
  clv = 0
  for (t in 0:(n_years-1)) {
    clv = clv + avgRev[t + 1]/((1+r)^t)
    # note we use avgRev[t+1] because avgRev values are indexed 1 to T, while t ranges from 0 to T-1
    }
  return(clv)
}


CLV_cohort_df<- rep(NA, length(all_segments))
total_value_cohort<-rep(NA, length(all_segments))
segment_size <- rep(NA, length(all_segments))


for (i in 1:length(all_segments)) {
```

```
  segment_DF <- subset(DF_comb, segment == i)
  profit_cohort<-rmargin*segment_DF[, 2:11]


  CLV_cohort_df[i]<-round(CLV_cohort(profit_cohort,return_rate), digits = 2)



    segment_size[i]<-nrow(segment_DF)
    total_value_cohort[i] <-CLV_cohort_df[i]*segment_size[i]



}



Answer_table_cohort<-round(data.frame(segment = all_segments,
                                      segment_size = segment_size,
                                final_averages_DF = final_averages_DF,
                                total_value = total_value_cohort,
                                CLV = CLV_cohort_df), digits = 2)

print(Answer_table_cohort)
```

```
##   segment segment_size final_averages_DF.spend_online0
## 1       4          428                           0.02
## 2       2          303                         110.37
## 3       1          102                          73.81
## 4       3          167                         188.38
##   final_averages_DF.spend_retail0 final_averages_DF.age final_averages_DF.white
## 1                           88.89                 41.56                    0.82
## 2                            0.03                 40.37                    0.79
## 3                           71.72                 40.84                    0.71
## 4                          195.40                 40.26                    0.82
##   final_averages_DF.college final_averages_DF.male final_averages_DF.hh_inc
## 1                      0.58                   0.00                   102.48
## 2                      0.48                   0.00                    85.08
## 3                      0.50                   0.89                    98.49
## 4                      0.58                   0.00                    99.22
##   total_value    CLV
## 1    54415.92 127.14
## 2    43038.12 142.04
## 3    22893.90 224.45
## 4    93003.97 556.91
```

Discussion:

- Which segments are expected to be most profitable on a per-customer basis? On a total value basis?

On per customer basis<- Segment 3> Segment 1> Segment 2> Segment 4 On total value basis<- Segment 3> Segment 4> Segment 2> Segment 1 Segment 3 has highest probability based on both per customer basis and total value basis.

- How similar are the CLV estimates from the simple and cohort methods?

The segment 4 seems to be most profitable followed by segment 3 then segment 2 and then 1.
On a total basis segment 4 is most profitable followed by segment 1 then segment 2 and then segment 3.

The cohort CLV estimates is lower by around 6.9% for segments 1&2 and around 14% for segment 4, for 10 year horzon however for segment 3, cohort is higher.