Workshop 8 – Solution

Customer retention and acquisition

MSBX-5130: Customer Analytics

3/12/2020

1) Objectives & setup

- Workshop tasks:
  1) Compute individual-level customer lifetime value (CLV) measures for a women's apparel brand
     - Estimate and anlayze a model of retention/churn (probability active)
     - Estimate and anlayze a model of profits given the customer is active
  2) Use results to assess profitability of customer retention, acquistion efforts
- The workshop makes use of the data file `customer_data.csv` – a panel dataset containing observations of total annual revenue for a sample of 1000 customers, over a period of 10 years.
  - Sampled customers are from the same "cohort", meaning they all became customers in the same year (0).
  - Demographic variables are included for all observations

The variables in the `customer_data.csv` are:

Variable	Description
`iid`	identifier for customer
`year`	identifier for year (0 = 1st year as customer)
`revenue`	total dollars spent by customer in year
`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)

Workshop task workflow

1. Setup
   1. Overview of approach to calculating individual-level CLV
   2. Download data & R Markdown files
   3. Load and summarize data file
2. Model of churn/retention
3. Model of profits given retention
4. Implementation of general CLV formula

5. Application to acqusition
    1. Prospective customer 1
    2. Prospective customer 2

1.1) Overivew of approach to calculating individual-level CLV

Today we explore methods to calculate CLV on an individual basis. The basic approach is to use models and historical data to predict:

a) individual retention rates over time (α_t), and possibly
b) individual revenues/profits for an *active* customer over time (m_t)

Both models use demographic variables (and other customer characteristics) and time (since first purchase) as predictors. Note that in contractual settings (e.g. cell phone service), model (b) may be unnecessary because profits are frequently fixed and known in advance – in this case we need only "plug in" (rather than predict) profit values for the CLV formula.

With estimates of retention rates (α_t) and profits given retention (m_t) over time for a given individual, we can evaluate the CLV (with horizon T and discount rate r) using the following "general" CLV formula:

$CLV = \sum_{t=0}^{T-1} \frac{\mathbb{E}[profit_t]}{(1+r)^t} = \sum_{t=0}^{T-1} \frac{\mathbb{E}[profit_t | \ active_t]Pr[active_t]}{(1+r)^t} = \sum_{t=0}^{T-1} \frac{m_t \alpha_t}{(1+r)^t}$

This type of analysis is particularly useful for targeting customer retention and acquisition activities. For example, using our estimated models, we can predict the expected CLV of new customers given the potential customer's demographic profile (or, as a proxy, the demographic profile of the geographic region they live in). Linking back to R, this would entail doing an "out of sample" prediction of retention rates (over the desired CLV horizon).

1.2) Download data & R Markdown file

If you have not already done so, download the data file `customer_data.csv` from Canvas. Also download this R markdown file, `Workshop8.Rmd`.

Now launch RStudio, and change the working directory to where you have downloaded the previously mentioned files.

1.3 Load and summarize data

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

```r
DF = read.csv('customer_data.csv')
head(DF)
```

```
  iid year revenue age     white   college male hh_inc
1  14    0  132.98  29 0.3241053 0.2868369    0 40.322
2  14    1  216.21  29 0.3241053 0.2868369    0 40.322
3  14    2  169.94  29 0.3241053 0.2868369    0 40.322
4  14    3   76.23  29 0.3241053 0.2868369    0 40.322
5  14    4  172.05  29 0.3241053 0.2868369    0 40.322
6  14    5    0.00  29 0.3241053 0.2868369    0 40.322
```

```r
summary(DF)
```

```
      iid             year        revenue             age       
 Min.   :   14   Min.   :0.0   Min.   :   0.00   Min.   :18.00  
 1st Qu.: 2946   1st Qu.:2.0   1st Qu.:   0.00   1st Qu.:33.00  
 Median : 5430   Median :4.5   Median :   0.00   Median :41.00  
 Mean   : 5463   Mean   :4.5   Mean   :  67.09   Mean   :40.91  
 3rd Qu.: 8110   3rd Qu.:7.0   3rd Qu.:  52.03   3rd Qu.:49.00  
 Max.   :10589   Max.   :9.0   Max.   :5456.27   Max.   :88.00  
     white           college            male           hh_inc       
 Min.   :0.0000   Min.   :0.0000   Min.   :0.000   Min.   :  2.499  
 1st Qu.:0.7297   1st Qu.:0.3835   1st Qu.:0.000   1st Qu.: 59.356  
 Median :0.8550   Median :0.5580   Median :0.000   Median : 87.364  
 Mean   :0.7993   Mean   :0.5437   Mean   :0.091   Mean   : 96.254  
 3rd Qu.:0.9422   3rd Qu.:0.7136   3rd Qu.:0.000   3rd Qu.:122.602  
 Max.   :1.0000   Max.   :1.0000   Max.   :1.000   Max.   :250.001  
```

2) Model of churn/retention

Before building our model, we first recall the relationship between retention and churn:

retention rate = $\alpha = Pr[active = 1]$

churn rate = 1 - α = 1 - $Pr[active = 1] = Pr[active = 0]$

This relationship implies that it is possible to estimate retention rates either from:

a) a predictive logit model with DV = `active` = (`revenue>0`) – retention form
b) a predictive logit model with DV = `inactive` = (`revenue==0`) – churn form

2.1) Retention form

Our objective here is to estimate a logit model, where each observation is associated with a customer/time period combination and the outcome ("active") is a binary indicator of whether or not the customer made a purchase in than period (i.e., if revenue > 0).

To estimate the model in retention form, perform the following tasks:

a) Create a new variable in the dataframe `DF` called `active`, which is equal to 1 if `DF$revenue` > 0 (and 0 otherwise)

b) Create a new dataframe `DF_active` using the subset of observations for which year > 0. We condition on years 1 and higher because by construction, year 0 retetion rates are 1. This further implies we will need to "manually" set the prediction of retention rates for year 0 to 1 before calling our CLV function.

c) Using `glm()` and data frame `DF_active`, estimate a model of retention by using `active` as the dependent variable and the following as independent variables: `year`, `year` squared, `age`, `white`, `college`, `male`, `hh_inc`. Please use this order for the independent variables in your call to `glm()`. Name the model `model_active`. Be sure to set `x=TRUE` as a parameter to `glm()`.

d) Use `summary()` to summarize the model results

```r
DF$active = as.numeric(DF$revenue>0)
DF_active = subset(DF,year>0)
#DF_retention = DF
model_active = glm(active ~ year + I(year^2) + age + white + college + male + hh_inc,
             data = DF_active,
             family = binomial(link = "logit"),
             x=TRUE)
summary(model_active)
```

```
Call:
glm(formula = active ~ year + I(year^2) + age + white + college +
    male + hh_inc, family = binomial(link = "logit"), data = DF_active,
    x = TRUE)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4934  -0.7792  -0.4751   0.8518   2.4267

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.2199140  0.1636957   1.343    0.179
year        -0.7907982  0.0464918 -17.009  < 2e-16 ***
I(year^2)    0.0354445  0.0046111   7.687 1.51e-14 ***
age          0.0233038  0.0021886  10.648  < 2e-16 ***
white       -0.0182896  0.1404681  -0.130    0.896
college      0.0703648  0.1504953   0.468    0.640
male        -0.3545611  0.0902212  -3.930 8.50e-05 ***
hh_inc       0.0111402  0.0006581  16.928  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 12000.0  on 8999  degrees of freedom
Residual deviance:  9347.3  on 8992  degrees of freedom
AIC: 9363.3

Number of Fisher Scoring iterations: 4
```

Discussion:

- Interpret the regression coefficients.

Intercept:

- *The intercept can be interpreted as utility when other regressors = 0*
- *The intercept can also be interpreted as the log-odds of "success" (customer being retained) when other regressors = 0*
  - *Utility, log-odds(active==1) = 0.220*

`year` and `year`^2:

- *The contribution of* `year` *to predicted utility/log-odds is -0.791x*`year` *+* `0.035`*x*`year`*^2. Increasing* `year` *by 1 unit increases utility/log-odds by -0.791 + 2x*`0.035`*x*`year`*, which depends on the value of* `year`*.*

age:

- *Each +1 year in age increases utility, and log odds of retention, by 0.023.*

Other continuous variables similar…

male:

- *Being male (*`male`*==1) decreases utility, and log odds of retention, by -0.35, compared to being female (*`male`*==0).*

2.2) Churn form

Our objective here is to estimate a logit model, where each observation is associated with a customer/time period combination and the outcome ("inactive") is a binary indicator of whether or not the customer made a purchase in than period (i.e., if revenue = 0).

To estimate the model in churn form, perform the following tasks:

a) Create a new variable in the dataframe `DF` called `inactive`, which is equal to 1 if DF$revenue == 0 (and 0 otherwise)

b) Create a new dataframe `DF_inactive` using the subset of observations for which year > 0. We condition on years 1 and higher because by construction, year 0 retetion rates are 1. This further implies we will need to "manually" set the prediction of retention rates for year 0 to 1 before calling our CLV function.

c) Using `glm()` and data frame `DF_inactive`, estimate a model of retention by using `inactive` as the dependent variable and the following as independent variables: `year`, `year` squared, `age`, `white`, `college`, `male`, `hh_inc`. Please use this order for the independent variables in your call to `glm()`. Name the model `model_inactive`. Be sure to set `x=TRUE` as a parameter to `glm()`.

d) Use `summary()` to summarize the model results

```
DF$inactive = as.numeric(DF$revenue==0)
DF_inactive = subset(DF,year>0)
model_inactive= glm(inactive ~  year + I(year^2) + age + white + college + male + hh_inc,
             data = DF_inactive,
             family = binomial(link = "logit"),
             x=TRUE)
summary(model_inactive)
```

```
Call:
glm(formula = inactive ~ year + I(year^2) + age + white + college +
    male + hh_inc, family = binomial(link = "logit"), data = DF_inactive,
    x = TRUE)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4267  -0.8518   0.4751   0.7792   2.4934
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.2199140  0.1636957  -1.343    0.179
year         0.7907982  0.0464918  17.009  < 2e-16 ***
I(year^2)   -0.0354445  0.0046111  -7.687 1.51e-14 ***
age         -0.0233038  0.0021886 -10.648  < 2e-16 ***
white        0.0182896  0.1404681   0.130    0.896
college     -0.0703648  0.1504953  -0.468    0.640
male         0.3545611  0.0902212   3.930 8.50e-05 ***
hh_inc      -0.0111402  0.0006581 -16.928  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 12000.0  on 8999  degrees of freedom
Residual deviance:  9347.3  on 8992  degrees of freedom
AIC: 9363.3

Number of Fisher Scoring iterations: 4
```

Discussion:

- What do you notice about the pattern of the coefficients between the "retention" form and the "churn" forms of the model? Why does this make sense?

The coefficients in the churn form have the same magnitude but opposite sign from the retention form. This makes sense because the dependent variables (active, inactive) are logical opposites of one another (e.g., active = 1 imples inactive=0).

2.3) Summary of demographic variables impact on churn

Here we use the churn form of the model to highlight which demographic varaibles drive customer churn. Specifically, we ignore time effects and focus on `age`, `white`, `college`, `male` and `hh_inc`. This section parallels churn management step 3 ("Use model to understand main drivers of churn") from the slides (17-19).

In the interest of time, I demonstrate how to compile the relevant data for a table comparable to that on slide 18. Recall that the main idea here is to infer which variables have the most influence in driving changes in the churn rate. To do this, we first compute the marginal effects, which give us the change in churn probability for a 1 unit change of a continuous regressor, or the change in churn probability for a binary varible (going from 0 to 1). Next, we estimate the change in probability by multiplying the marginal effect by the standard unit of variablility for the associated variable (standard deviation for continuous variables, 1 for binary variables).

When you get to this point in the workshop, uncomment lines beginning with ## below.

```r
# NOTE: uncomment lines beginning with ## when you get here
# calculate marginal effects
library(erer)
model_inactive.me = maBina(model_inactive, x.mean = FALSE, digits = 3)

# calculate delta X for marginal effects
```

```r
# delta X = sd(var) if var is continuous, 1 if var is binary
stddev = rep(0,5)
stddev[1] = sd(DF$age)
stddev[2] = sd(DF$white)
stddev[3] = sd(DF$college)
stddev[4] = 1                # male is a binary variable
stddev[5] = sd(DF$hh_inc)

# create table of coefficients, z stats, marginal effects, delta X, delta prob
# note that in model specification, demographic variables are indexed by 4:8
DF_disp = data.frame(coef=coef(summary(model_inactive))[4:8, 1],
                     z_stat=coef(summary(model_inactive))[4:8, 3],
                     marginal_eff=model_inactive.me$out[4:8, 1],
                     delta_X=stddev)
DF_disp$delta_prob = DF_disp$marginal_eff*DF_disp$delta_X

# display table
library(knitr)
kable(DF_disp,digits=3)
```

	coef	z_stat	marginal_eff	delta_X	delta_prob
age	-0.023	-10.648	-0.004	11.802	-0.047
white	0.018	0.130	0.003	0.196	0.001
college	-0.070	-0.468	-0.012	0.217	-0.003
male	0.355	3.930	0.077	1.000	0.077
hh_inc	-0.011	-16.928	-0.002	50.157	-0.100

Discussion:

- Which demographic variables have the greatest influence on churn rates? Which may be safely ignored?

Income has the largest magnitude (negative) effect, followed by male (positive), followed by age (negative). Both white and college are insignigicant and can thus be ignored.

- Based on these insights, suggest a strategy for the firm improve future retention rates.

The firm cannot affect the income, age or gender of their current customers. However, the firm can target affluent older women as new customers. To the extent the firm can increase the proportion of such customers in the customer base, retention rates will improve.

3) Model of profits given retention

Recall that in contractual settings, such as cell phone service or app subscriptions, if the customer maintains the contract/service, revenues are known in advance (e.g. $70/month for cell service). Since margins and operating costs are relatively stanble in such environments, profits can usually be approximated directly with adjustments to revenue (using knowledge of operating margins and/or any recurring fixed costs). In contractual settings, we can therefore generally forecast profits (given customer is active) without the use of a further model.

In this case, we have data from a non-contractual setting. We know this from the data description and from inspecting the revenue data, which is clearly variable over time. In this case, we need a separate model of profits conditional upon the customer being active.

A simple approach to a predictive model of profits conditional upon the customer being active is:

1. Create a dataframe that conditions only on observations where customers are active
2. Use `lm()` to estimate a model of customer revenues as a function of time and demographic variables
3. Adjust revenue predictions from `lm()` model as needed to convert to profits
    - Predictions may be for existing customers (in-sample) or prospective customers (out of sample)

Following steps 1 and 2 above, use `lm()` and data frame `DF` to estimate a model of revenues by year using **`revenue`** as the dependent variable and the following as independent variables: **`year`**, **`year`** squared, **`age`**, **`white`**, **`college`**, **`male`**, **`hh_inc`**. Name the model `model_rev`. Use `summary()` to summarize the model results.

```
DF_rev = subset(DF,active==1)
model_rev = lm(revenue~year+I(year^2)+age+white+college+male+hh_inc,
               data=DF_rev)
summary(model_rev)
```

```
Call:
lm(formula = revenue ~ year + I(year^2) + age + white + college +
    male + hh_inc, data = DF_rev)

Residuals:
   Min     1Q Median     3Q    Max
-231.2 -120.2  -78.8    0.7 5226.5

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   73.72566   24.22245   3.044 0.002351 **
year          -3.74484    5.12211  -0.731 0.464748
I(year^2)      0.04182    0.64434   0.065 0.948254
age            1.24099    0.35148   3.531 0.000419 ***
white         53.22672   24.08001   2.210 0.027127 *
college     -104.09768   25.20117  -4.131 3.68e-05 ***
male          -4.80481   15.16854  -0.317 0.751439
hh_inc         0.47143    0.09995   4.717 2.47e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 278.7 on 4461 degrees of freedom
Multiple R-squared:  0.01079,   Adjusted R-squared:  0.009242
F-statistic: 6.954 on 7 and 4461 DF,  p-value: 2.919e-08
```

3.1) Summary of demographic variables impact on revenue | active

Following the example in section 2.3, create a table summarizing the impact of the demographic variables on revenues (given the customer is active).

Hint: Recall that for a linear regression, the marginal effect is equal to the parameter estimate.

```
DF_disp = data.frame(coef=coef(summary(model_rev))[4:8, 1],
                     z_stat=coef(summary(model_rev))[4:8, 3],
                     marginal_eff=coef(summary(model_rev))[4:8, 1],
                     delta_X=stddev)
DF_disp$delta_rev = DF_disp$marginal_eff*DF_disp$delta_X

# display table
library(knitr)
kable(DF_disp,digits=3)
```

	coef	z_stat	marginal_eff	delta_X	delta_rev
age	1.241	3.531	1.241	11.802	14.646
white	53.227	2.210	53.227	0.196	10.423
college	-104.098	-4.131	-104.098	0.217	-22.612
male	-4.805	-0.317	-4.805	1.000	-4.805
hh_inc	0.471	4.717	0.471	50.157	23.646

Discussion:

- Which demographic variables have the greatest influence on revenues (given active)? Which may be safely ignored?

Income has the largest magnitude (positive) effect, followed by college (negative), followed by age (positive), followed by white (positive). Male is insignigicant and can thus be ignored.

4) Implementation of general CLV formula

At this point, we have models estimated for the retention rate and for revenues given retention. Before using the models to predict CLV, we first implement a CLV formula that will accept a list of retention rates over time and profits given retention over time.

Recall that a general formula for CLV is $CLV = \sum_{t=0}^{T-1} \frac{m_t * \alpha_t}{(1+r)^t}$, where m_t and α_t are the profit (given retention) and the retention rate in period t, respectively, and r is the discount (interest) rate. So, we design the function to accept two lists or vectors of (equal) length. The first list/vector represents the per-period average profits (given retention) as a function of time, while the second list/vector represents the retention rate over time. Note that we will assume the first value of each list is associated with period 0, which implies: a) the number of future periods modeled will be the length of the list minus one, and b) $\alpha_0 = 1$, because all customers are present in period 0 by definition.

To sum up, write the function to accept as input: a) sequence of profit values (from period 0 to T-1), b) sequence of retention rates (alphas, from period 0 to T-1), c) discount rate (r). The function should return the computed CLV value. Name the function `CLV_general()`. Hint: As with the cohort method, take care to properly index time and the elements of the inputs.

Next, demonstrate the use of your function. For this, we will assume T=6, meaning the length of the lists submitted to `CLV_general()` will be 6. Specifically, assume: a) profits = 100, b) retention rates = list of length 6, beginning with 1 and declining by 4% per period (e.g. 1, .96, .92, . . .), c) interest rate = 10%.

```
# CLV_general(M,alpha,r)
# M = list of profits per period, indexed 0 to T-1
```

```
# alpha = list of retention rates per period, indexed 0 to T-1
# r = discount rate [0,1]
CLV_general = function(M,alpha,r) {
  T = length(M)
  clv = 0
  for (t in 0:(T-1)) {
    clv = clv + M[t+1]*alpha[t+1]/((1+r)^t)
    }
  out = as.numeric(clv)
  }

M = rep(100,6)
alpha = seq(from=1,by=-.04,length.out=6)
r = .1
clv3 = CLV_general(M,alpha,r); clv3
```

```
[1] 436.4683
```

5 Application to acqusition

We demonstrate application of the model by evaluating two prospective customers in terms of expected CLV.

5.1) Prospective customer 1

Using the results from `model_active` and `model_revenue`, predict CLV over a 6 year period (years 0 to 5), for a customer with the following demographic profile: a) age = 20, b) white = 1, college = 0, male = 0 hh_inc = 30. Further assume the retailer margin is still 40% and the interest rate r = 10%.

To do this, perform the following steps:

a) Create a data frame, called `DF_eval1`, with 5 rows and the following columns: year, age, white, college, male, hh_inc. Year should range from 1 to 5, while all other variables are constant over time (rows). As a hint, recall that you can easily build data frames from lists as follows:

```
year = c(1,2,3)
age = rep(20,3)
DF_tmp = data.frame(year,age)
DF_tmp
```

```
  year age
1    1  20
2    2  20
3    3  20
```

b) Using `predict()` or equivalent methods, predict the retention probabilties for this customer over years 1 to 5. Store the predictions in a list called `p_active`.

c) Construct the complete list of retention probabilities. The first element of the list should be 1 (corresponding to period 0), while the remaining elements are those of `p_active`. Name the resulting list (length 6) `alpha1`. Print the list `alpha1`.

d) Create a data frame, called `DF_eval2`, with 6 rows and the following columns: year, age, white, college, male, hh_inc. Year should range from 0 to 5, while all other variables are constant over time (rows).

e) Using `predict()` or equivalent methods, predict revenues (given retention) for this customer over years 0 to 5. Multiply revenues by the margin (40%) to get profits. Store the profit predictions in a list called `M1`. Print the list `M1`.

f) Use your function `CLV_general()` to compute the expected CLV for this customer, taking as inputs `M1` and `alpha1` and assuming the interest rate r=.1 (10%). Print the result.

```
year = c(1,2,3,4,5)
age = rep(20,5)
white = rep(1,5)
college = rep(0,5)
male = rep(0,5)
hh_inc = rep(30,5)
DF_eval1 = data.frame(year,age,white,college,male,hh_inc)
p_active = predict(model_active, newdata=DF_eval1, type = "response")
alpha1 = c(1,p_active); alpha1
```

```
                  1         2         3         4         5
1.0000000 0.5613278 0.3922386 0.2589395 0.1687974 0.1124484
```

```
year = c(0,1,2,3,4,5)
age = rep(20,6)
white = rep(1,6)
college = rep(0,6)
male = rep(0,6)
hh_inc = rep(30,6)
DF_eval2 = data.frame(year,age,white,college,male,hh_inc)
rev_hat = predict(model_rev, newdata=DF_eval2); rev_hat
```

```
       1        2        3        4        5        6
165.9151 162.2121 158.5927 155.0570 151.6049 148.2364
```

```
M1 = .4*rev_hat; M1
```

```
       1        2        3        4        5        6
66.36606 64.88485 63.43710 62.02280 60.64196 59.29457
```

```
clv5a = CLV_general(M1,alpha1,r); clv5a
```

```
[1] 143.2384
```

Discussion:

- What is the maximum acquistion cost you would be willing to pay for this customer, assuming you want to break-even within 6 years?

The CLV of $143.24 is the maximum we should pay if we want to break even on this customer within 6 years.

5.2) Prospective customer 2

Repeat the exercise in the previous section, for a customer with the following demographic profile: a) age = 50, b) white = 1, college = 1, male = 0 hh_inc = 100. Assume that the retailer margin is still 40%. Print the predicted retention rates, profits given retention, and the computed CLV value.

```
year = c(1,2,3,4,5)
age = rep(50,5)
white = rep(1,5)
college = rep(1,5)
male = rep(0,5)
hh_inc = rep(100,5)
DF_eval1 = data.frame(year,age,white,college,male,hh_inc)
p_active = predict(model_active, newdata=DF_eval1, type = "response")
alpha2 = c(1,p_active); alpha2
```

```
                  1         2         3         4         5
1.0000000 0.8576425 0.7523865 0.6219435 0.4887816 0.3736291
```

```
year = c(0,1,2,3,4,5)
age = rep(50,6)
white = rep(1,6)
college = rep(1,6)
male = rep(0,6)
hh_inc = rep(100,6)
DF_eval2 = data.frame(year,age,white,college,male,hh_inc)
rev_hat = predict(model_rev, newdata=DF_eval2); rev_hat
```

```
       1        2        3        4        5        6
132.0474 128.3444 124.7250 121.1893 117.7372 114.3687
```

```
M2 = .4*rev_hat; M2
```

```
       1        2        3        4        5        6
52.81898 51.33777 49.89001 48.47572 47.09487 45.74749
```

```
clv5b = CLV_general(M2,alpha2,r); clv5b
```

```
[1] 172.8547
```

Discussion:

- Under the given assumptions and data, which prospective customer is expected to be more profitable?

Prospective customer 2 is expected to be more profitable.