Data Analysis Script on Predictive model selection for completion rate in Massive Open Online Courses

A. De Santis,* *K. Sannicandro, C. Bellini, T. Minerva*

Abstract

This RMD file contains the script we used to perform the data analysis described in the paper. More detailed statistical results and comments are included in this addendum

Contents

1 Descriptive Analysis

- Purged data removing N/A: 722 observations;

---
*University of Modena and Reggio Emilia, annamaria.desantis@unimore.it

Table 1: Certificate download Frequency Table

	Frequencies	% Frequencies
Not Completed	473	65.51
Completed w/Certificate Download	249	34.49
TOTAL	**722**	**100.00**

- Reversed values for CERTIFICATE, MARRIED, CHILDREN, CTUTORED
- In ZIP File: data files, rmd
- just change the variable DIRECTORY in the RMD script to reproduce
- Use R-Studio with Knitr to reproduce

1.1 Certificate Downloads

1.1.1 Frequency Distribution

To be noticed: 34,5% of users completed the course and downloaded the certificate.

Also explain that this rate can under-estimate the rate of users who completed the course or, at least, who completed all the learning activities.

1.1.2 Pie-Chart showing the different rates
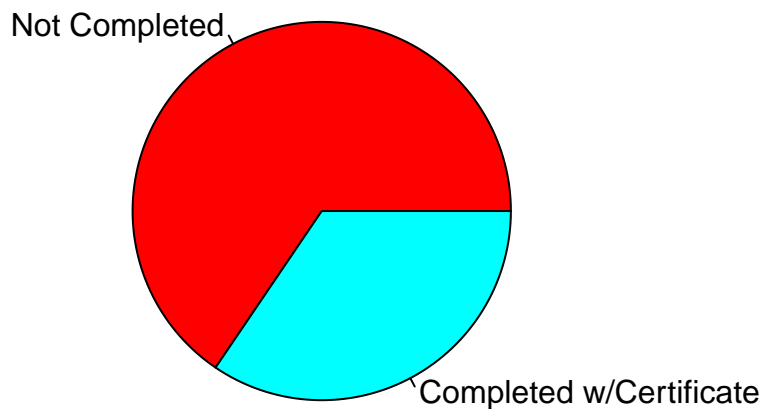
1.2 Completion Rate

1.2.1 Frequency Distribution

To be noticed:

- more than 44% of user completed at least 90% of learning activities. We can consider them as 'Completed'.
- about 39% of users cover less then 20% af learning activities;
- from 20% to 90% of completion rate we have about 17% of users. They could be the users who were engaged in the course but adandoned it.. no reason why.

1.2.2 Histogram of Completion Rate

To be noticed:

Table 2: Completion Rate Frequency Table

	Frequencies	% Frequencies
0 to 10%	208	28.81
10+ to 20%	73	10.11
20+ to 30%	26	3.60
30+ to 40%	34	4.71
40+ to 50%	25	3.46
50+ to 60%	12	1.66
60+ to 70%	4	0.55
70+ to 80%	3	0.42
80+ to 90%	17	2.35
90+ to 100%	320	44.32
TOTAL	**722**	**100.00**

- nearly bimodal distribution;
- modal bin > 90%-100%
- users who completed at least 50% tend to complete the course;

Histogram of Course Completion Rate
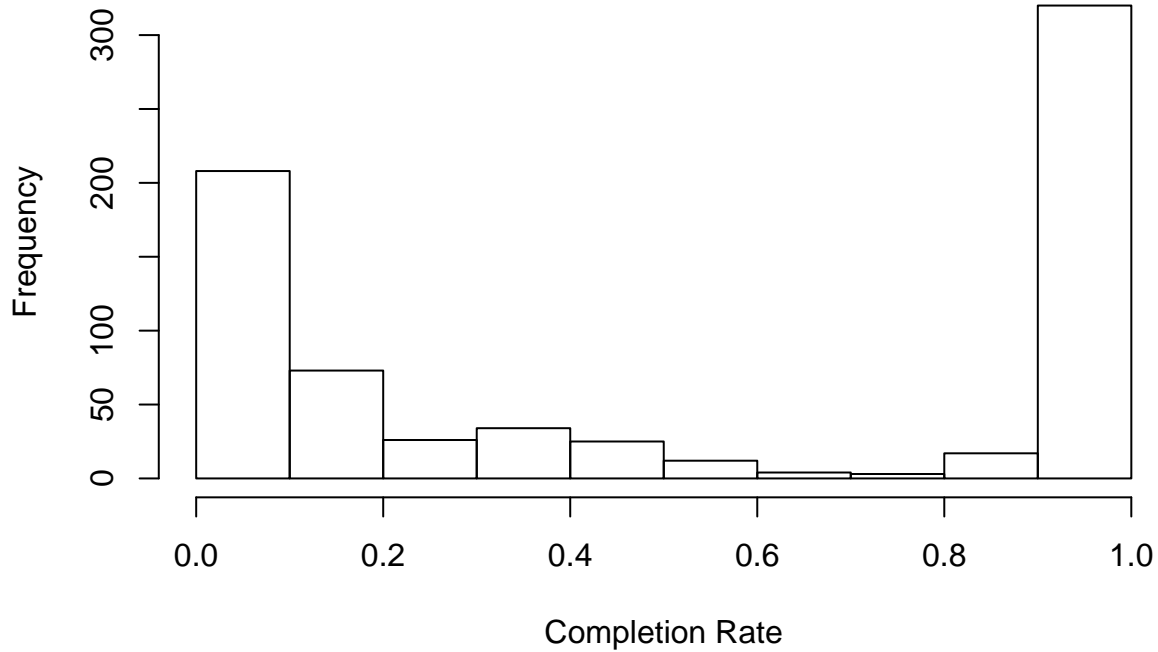
2   Correlation Analysis

2.1   Variables Blocks

We set 5 blocks of variables:

- Response variables: Completion Rate and Certificate Downloads
- User Profile
- Course Profile

- User Engagement
- User Behavior

2.2 Intra-Blocks Correlation

	CRATE	CERTIFICATE
CRATE	1.00	0.77
CERTIFICATE	0.77	1.00

	GENDER	DEGREE	LANGUAGE	AGE	MARRIED	CHILDREN	TRAINING	WORKING	SECTOR	DIGITAL
GENDER	1.00	0.08	0.00	-0.11	-0.13	-0.15	0.08	0.07	0.04	-0.09
DEGREE	0.08	1.00	0.10	0.26	0.13	0.08	0.13	-0.11	-0.10	0.08
LANGUAGE	0.00	0.10	1.00	-0.10	-0.04	-0.08	0.02	-0.04	0.10	0.12
AGE	-0.11	0.26	-0.10	1.00	0.50	0.51	0.01	-0.33	-0.34	-0.03
MARRIED	-0.13	0.13	-0.04	0.50	1.00	0.64	-0.07	-0.18	-0.17	0.05
CHILDREN	-0.15	0.08	-0.08	0.51	0.64	1.00	-0.07	-0.21	-0.16	0.04
TRAINING	0.08	0.13	0.02	0.01	-0.07	-0.07	1.00	0.08	0.06	0.01
WORKING	0.07	-0.11	-0.04	-0.33	-0.18	-0.21	0.08	1.00	0.21	0.02
SECTOR	0.04	-0.10	0.10	-0.34	-0.17	-0.16	0.06	0.21	1.00	-0.05
DIGITAL	-0.09	0.08	0.12	-0.03	0.05	0.04	0.01	0.02	-0.05	1.00

	EFFORT	PRE.KNOWLEDGE	DROPOUT_TOT	DROPOUT_INT	DROPOUT_LEA	DROPOUT_NAV	MOTIVATION
EFFORT	1.00	0.11	-0.03	0.04	-0.02	-0.03	0.03
PRE.KNOWLEDGE	0.11	1.00	0.10	0.18	0.08	0.03	0.32
DROPOUT_TOT	-0.03	0.10	1.00	0.68	0.91	0.76	0.26
DROPOUT_INT	0.04	0.18	0.68	1.00	0.51	0.33	0.34
DROPOUT_LEA	-0.02	0.08	0.91	0.51	1.00	0.54	0.19
DROPOUT_NAV	-0.03	0.03	0.76	0.33	0.54	1.00	0.18
MOTIVATION	0.03	0.32	0.26	0.34	0.19	0.18	1.00

	CTUTORED	CCAT	CLANG	CHOUR	CLEVEL
CTUTORED	1.00	-0.57	0.57	0.34	0.57
CCAT	-0.57	1.00	-1.00	-0.49	-1.00
CLANG	0.57	-1.00	1.00	0.49	1.00
CHOUR	0.34	-0.49	0.49	1.00	0.49
CLEVEL	0.57	-1.00	1.00	0.49	1.00

	CLICK_MANDATORY	CLIC_TOTAL
CLICK_MANDATORY	1.00	0.92
CLIC_TOTAL	0.92	1.00

2.3 Response vs Predictors Blocks correlation

	CRATE	CERTIFICATE
GENDER	-0.13	-0.20
DEGREE	-0.01	-0.03
LANGUAGE	-0.09	-0.13
AGE	0.12	0.16
MARRIED	0.04	0.07
CHILDREN	0.10	0.11
TRAINING	0.04	0.04
WORKING	-0.01	-0.02
SECTOR	-0.02	0.00
DIGITAL	0.01	0.03

	CRATE	CERTIFICATE
CLICK_MANDATORY	0.77	0.64
CLIC_TOTAL	0.86	0.73

	CRATE	CERTIFICATE
CTUTORED	0.02	-0.08
CCAT	0.03	0.04
CLANG	-0.03	-0.04
CHOUR	0.03	-0.05
CLEVEL	-0.03	-0.04

	CRATE	CERTIFICATE
EFFORT	0.03	0.00
PRE.KNOWLEDGE	0.09	0.09
DROPOUT_TOT	-0.02	0.04
DROPOUT_INT	0.04	0.07
DROPOUT_LEA	-0.02	0.02
DROPOUT_NAV	0.02	0.05
MOTIVATION	0.17	0.15

To be noticed: check the R2 high values and comment

3 Stepwise Analysis

We performed a full stepwise analysis in the two cases where response variable is Completion Rate or Certificate Download

- AIC as selection parameter
- Stepwise method (back and forward)
- Output: Selected Model, Coefficients, R2, MSE, p for each variable

3.1 Certificate Downloads

```
##
## Call:
## lm(formula = CERTIFICATE ~ GENDER + LANGUAGE + DIGITAL + DROPOUT_TOT +
##     CLIC_TOTAL + CTUTORED + CCAT, data = General_C)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.17351 -0.15809 -0.00354  0.18856  0.75842
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.083930   0.069029  -1.216   0.2244
## GENDER      -0.111571   0.024365  -4.579 5.51e-06 ***
## LANGUAGE    -0.094688   0.039962  -2.369   0.0181 *
## DIGITAL      0.026646   0.012875   2.070   0.0389 *
## DROPOUT_TOT  0.003888   0.001533   2.536   0.0114 *
## CLIC_TOTAL   0.151043   0.005124  29.477  < 2e-16 ***
## CTUTORED    -0.124468   0.030014  -4.147 3.77e-05 ***
## CCAT        -0.060952   0.029962  -2.034   0.0423 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.311 on 714 degrees of freedom
## Multiple R-squared:  0.5767, Adjusted R-squared:  0.5726
## F-statistic:   139 on 7 and 714 DF,  p-value: < 2.2e-16
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.084	0.069	-1.216	0.224
GENDER	-0.112	0.024	-4.579	0.000
LANGUAGE	-0.095	0.040	-2.369	0.018
DIGITAL	0.027	0.013	2.070	0.039
DROPOUT_TOT	0.004	0.002	2.536	0.011
CLIC_TOTAL	0.151	0.005	29.477	0.000
CTUTORED	-0.124	0.030	-4.147	0.000
CCAT	-0.061	0.030	-2.034	0.042

3.2 Completion Rate

```
##
## Call:
## lm(formula = CRATE ~ GENDER + DEGREE + AGE + CHILDREN + SECTOR +
##     EFFORT + DROPOUT_TOT + DROPOUT_INT + DROPOUT_LEA + DROPOUT_NAV +
##     MOTIVATION + CLICK_MANDATORY + CLIC_TOTAL + CTUTORED + CCAT +
##     CHOUR, data = General_R)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.63249 -0.12630 -0.03239  0.15515  0.54720
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.2466447 0.0851396  -2.897  0.00388 **
```

```
## GENDER          -0.0490253  0.0174840  -2.804  0.00519 **
## DEGREE           0.0128663  0.0062292   2.065  0.03924 *
## AGE             -0.0108788  0.0044479  -2.446  0.01470 *
## CHILDREN         0.0311557  0.0205749   1.514  0.13041
## SECTOR          -0.0033587  0.0022009  -1.526  0.12746
## EFFORT           0.0005178  0.0003020   1.715  0.08685 .
## DROPOUT_TOT     -0.0127586  0.0068399  -1.865  0.06255 .
## DROPOUT_INT      0.0121188  0.0080558   1.504  0.13293
## DROPOUT_LEA      0.0137556  0.0080912   1.700  0.08956 .
## DROPOUT_NAV      0.0123300  0.0079863   1.544  0.12306
## MOTIVATION       0.0027059  0.0013242   2.043  0.04139 *
## CLICK_MANDATORY -0.0424384  0.0088380  -4.802 1.92e-06 ***
## CLIC_TOTAL       0.2197929  0.0129910  16.919  < 2e-16 ***
## CTUTORED         0.0642040  0.0255033   2.517  0.01204 *
## CCAT             0.1263252  0.0294622   4.288 2.06e-05 ***
## CHOUR            0.0163307  0.0033336   4.899 1.20e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2165 on 705 degrees of freedom
## Multiple R-squared:  0.756,  Adjusted R-squared:  0.7504
## F-statistic: 136.5 on 16 and 705 DF,  p-value: < 2.2e-16
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.247	0.085	-2.897	0.004
GENDER	-0.049	0.017	-2.804	0.005
DEGREE	0.013	0.006	2.065	0.039
AGE	-0.011	0.004	-2.446	0.015
CHILDREN	0.031	0.021	1.514	0.130
SECTOR	-0.003	0.002	-1.526	0.127
EFFORT	0.001	0.000	1.715	0.087
DROPOUT_TOT	-0.013	0.007	-1.865	0.063
DROPOUT_INT	0.012	0.008	1.504	0.133
DROPOUT_LEA	0.014	0.008	1.700	0.090
DROPOUT_NAV	0.012	0.008	1.544	0.123
MOTIVATION	0.003	0.001	2.043	0.041
CLICK_MANDATORY	-0.042	0.009	-4.802	0.000
CLIC_TOTAL	0.220	0.013	16.919	0.000
CTUTORED	0.064	0.026	2.517	0.012
CCAT	0.126	0.029	4.288	0.000
CHOUR	0.016	0.003	4.899	0.000

4 Residual Analysis

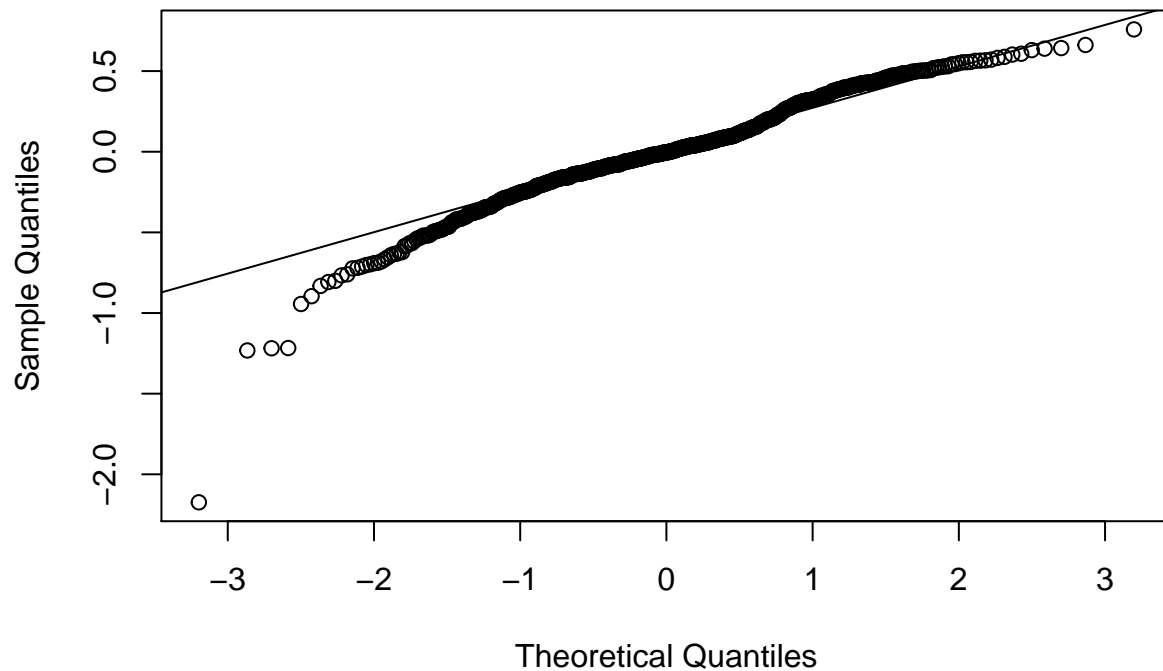## 4.1 Certificate Downloads

```
##
##  Shapiro-Wilk normality test
##
## data:  resC
## W = 0.95758, p-value = 1.38e-13

##
```

```
##  Anderson-Darling normality test
##
## data:  resC
## A = 4.2802, p-value = 1.221e-10

##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  resC
## D = 0.065701, p-value = 7.987e-08
```
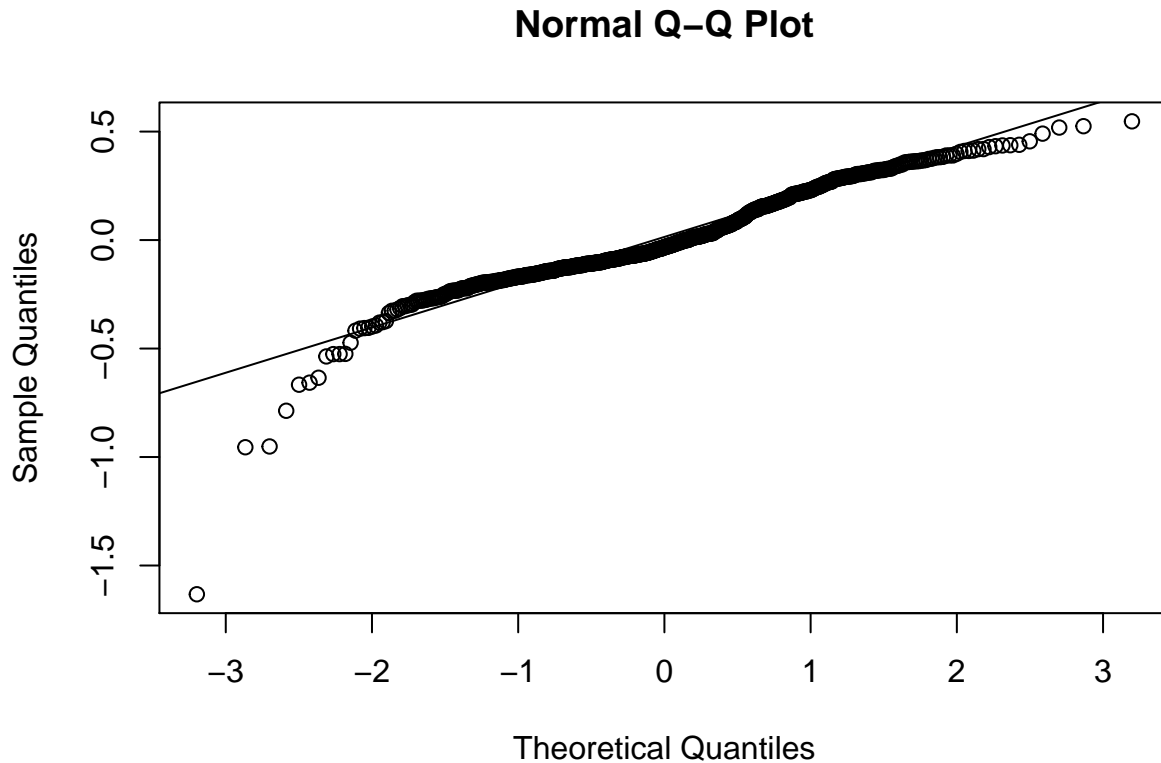
Normal Q–Q Plot

4.2   Completion Rate

```
##
##  Shapiro-Wilk normality test
##
## data:  resR
## W = 0.93921, p-value < 2.2e-16

##
##  Anderson-Darling normality test
##
## data:  resR
## A = 7.3101, p-value < 2.2e-16

##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  resR
## D = 0.072751, p-value = 1.064e-09
```

8

Normal Q–Q Plot

5 Block Stepwise Analysis (Target vs Profile + Course)

5.1 Certificate Downloads

```
##
## Call:
## lm(formula = CERTIFICATE ~ GENDER + LANGUAGE + AGE + SECTOR,
##     data = General_C)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6290 -0.3594 -0.2283  0.5108  0.9789
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.302203   0.050396    5.997 3.19e-09 ***
## GENDER      -0.178139   0.034485   -5.166 3.11e-07 ***
## LANGUAGE    -0.191186   0.058419   -3.273  0.00112 **
## AGE          0.031163   0.007675    4.060 5.45e-05 ***
## SECTOR       0.008614   0.004627    1.862  0.06306 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.458 on 717 degrees of freedom
## Multiple R-squared:  0.078,  Adjusted R-squared:  0.07286
## F-statistic: 15.16 on 4 and 717 DF,  p-value: 6.576e-12
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.302	0.050	5.997	0.000
GENDER	-0.178	0.034	-5.166	0.000
LANGUAGE	-0.191	0.058	-3.273	0.001
AGE	0.031	0.008	4.060	0.000
SECTOR	0.009	0.005	1.862	0.063

5.2 Completion Rate

```
##
## Call:
## lm(formula = CRATE ~ GENDER + LANGUAGE + AGE + MARRIED + CHILDREN +
##     TRAINING + CTUTORED + CCAT, data = General_R)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.81994 -0.41426 -0.00843  0.40362  0.72666
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.416701   0.054329   7.670 5.66e-14 ***
## GENDER      -0.134171   0.033230  -4.038 5.98e-05 ***
## LANGUAGE    -0.101668   0.054124  -1.878  0.06073 .
## AGE          0.021893   0.008112   2.699  0.00712 **
## MARRIED     -0.071460   0.043037  -1.660  0.09727 .
## CHILDREN     0.067696   0.046357   1.460  0.14464
## TRAINING     0.009938   0.006535   1.521  0.12878
## CTUTORED     0.120154   0.041204   2.916  0.00366 **
## CCAT         0.092103   0.040902   2.252  0.02464 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4236 on 713 degrees of freedom
## Multiple R-squared:  0.05494,    Adjusted R-squared:  0.04433
## F-statistic: 5.181 on 8 and 713 DF,  p-value: 2.635e-06
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.417	0.054	7.670	0.000
GENDER	-0.134	0.033	-4.038	0.000
LANGUAGE	-0.102	0.054	-1.878	0.061
AGE	0.022	0.008	2.699	0.007
MARRIED	-0.071	0.043	-1.660	0.097
CHILDREN	0.068	0.046	1.460	0.145
TRAINING	0.010	0.007	1.521	0.129
CTUTORED	0.120	0.041	2.916	0.004
CCAT	0.092	0.041	2.252	0.025

6 Block Stepwise Analysis (Target vs Engagement + Behavior)

6.1 Certificate Downloads

```
##
```

```
## Call:
## lm(formula = CERTIFICATE ~ CLIC_TOTAL + CLICK_MANDATORY + PRE.KNOWLEDGE +
##     DROPOUT_TOT, data = General_C)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.21023 -0.14266 -0.00932  0.20764  0.69634
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -0.204726   0.052003  -3.937 9.06e-05 ***
## CLIC_TOTAL       0.198794   0.013593  14.625  < 2e-16 ***
## CLICK_MANDATORY -0.033292   0.008887  -3.746 0.000194 ***
## PRE.KNOWLEDGE    0.023457   0.011484   2.043 0.041465 *
## DROPOUT_TOT      0.003460   0.001581   2.188 0.029026 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3195 on 717 degrees of freedom
## Multiple R-squared:  0.5514, Adjusted R-squared:  0.5489
## F-statistic: 220.3 on 4 and 717 DF,  p-value: < 2.2e-16
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.205	0.052	-3.937	0.000
CLIC_TOTAL	0.199	0.014	14.625	0.000
CLICK_MANDATORY	-0.033	0.009	-3.746	0.000
PRE.KNOWLEDGE	0.023	0.011	2.043	0.041
DROPOUT_TOT	0.003	0.002	2.188	0.029

6.2 Completion Rate

```
##
## Call:
## lm(formula = CRATE ~ CLIC_TOTAL + CLICK_MANDATORY + PRE.KNOWLEDGE +
##     DROPOUT_TOT + DROPOUT_INT + DROPOUT_LEA + DROPOUT_NAV + MOTIVATION,
##     data = General_R)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.50218 -0.13549 -0.05537  0.14772  0.61834
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.066090   0.042876   1.541   0.1237
## CLIC_TOTAL       0.181588   0.009601  18.914   <2e-16 ***
## CLICK_MANDATORY -0.015752   0.006242  -2.524   0.0118 *
## PRE.KNOWLEDGE    0.014642   0.008433   1.736   0.0829 .
## DROPOUT_TOT     -0.015467   0.006915  -2.237   0.0256 *
## DROPOUT_INT      0.016440   0.008064   2.039   0.0419 *
## DROPOUT_LEA      0.017666   0.008205   2.153   0.0317 *
## DROPOUT_NAV      0.012436   0.008118   1.532   0.1260
## MOTIVATION       0.002717   0.001402   1.938   0.0530 .
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2224 on 713 degrees of freedom
## Multiple R-squared:  0.7396, Adjusted R-squared:  0.7367
## F-statistic: 253.2 on 8 and 713 DF,  p-value: < 2.2e-16
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.066	0.043	1.541	0.124
CLIC_TOTAL	0.182	0.010	18.914	0.000
CLICK_MANDATORY	-0.016	0.006	-2.524	0.012
PRE.KNOWLEDGE	0.015	0.008	1.736	0.083
DROPOUT_TOT	-0.015	0.007	-2.237	0.026
DROPOUT_INT	0.016	0.008	2.039	0.042
DROPOUT_LEA	0.018	0.008	2.153	0.032
DROPOUT_NAV	0.012	0.008	1.532	0.126
MOTIVATION	0.003	0.001	1.938	0.053

7 Block Stepwise Analysis (Target vs Behavior)

7.1 Certificate Downloads

```
##
## Call:
## lm(formula = CERTIFICATE ~ CLIC_TOTAL + CLICK_MANDATORY, data = General_C)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.19758 -0.13307 -0.00196  0.20284  0.68963
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -0.05100    0.01803  -2.828 0.004809 **
## CLIC_TOTAL       0.19782    0.01365  14.490  < 2e-16 ***
## CLICK_MANDATORY -0.03247    0.00893  -3.637 0.000296 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3213 on 719 degrees of freedom
## Multiple R-squared:  0.5451, Adjusted R-squared:  0.5439
## F-statistic: 430.8 on 2 and 719 DF,  p-value: < 2.2e-16
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.051	0.018	-2.828	0.005
CLIC_TOTAL	0.198	0.014	14.490	0.000
CLICK_MANDATORY	-0.032	0.009	-3.637	0.000

7.2 Completion Rate

```
##
## Call:
## lm(formula = CRATE ~ CLIC_TOTAL + CLICK_MANDATORY, data = General_R)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.50817 -0.13998 -0.06552  0.15912  0.62077
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.119872   0.012576   9.532   <2e-16 ***
## CLIC_TOTAL      0.183712   0.009522  19.293   <2e-16 ***
## CLICK_MANDATORY -0.015689   0.006228  -2.519    0.012 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2241 on 719 degrees of freedom
## Multiple R-squared:  0.7334, Adjusted R-squared:  0.7327
## F-statistic: 988.9 on 2 and 719 DF,  p-value: < 2.2e-16
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.120	0.013	9.532	0.000
CLIC_TOTAL	0.184	0.010	19.293	0.000
CLICK_MANDATORY	-0.016	0.006	-2.519	0.012

8   Block Stepwise Analysis (Target vs Engagement)

8.1   Certificate Downloads

```
##
## Call:
## lm(formula = CERTIFICATE ~ MOTIVATION, data = General_C)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5271 -0.3552 -0.2907  0.5911  0.8490
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.097296   0.063226   1.539    0.124
## MOTIVATION  0.010744   0.002636   4.075 5.11e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4706 on 720 degrees of freedom
## Multiple R-squared:  0.02255,   Adjusted R-squared:  0.02119
## F-statistic: 16.61 on 1 and 720 DF,  p-value: 5.108e-05
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.097	0.063	1.539	0.124
MOTIVATION	0.011	0.003	4.075	0.000

8.2   Completion Rate

```
##
## Call:
## lm(formula = CRATE ~ DROPOUT_TOT + DROPOUT_INT + DROPOUT_LEA +
```

```
##     DROPOUT_NAV + MOTIVATION, data = General_R)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -0.77467 -0.41537 -0.02228  0.41368  0.72374
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.448482   0.075191   5.965 3.85e-09 ***
## DROPOUT_TOT -0.049335   0.013068  -3.775 0.000173 ***
## DROPOUT_INT  0.047313   0.015263   3.100 0.002012 **
## DROPOUT_LEA  0.047207   0.015559   3.034 0.002501 **
## DROPOUT_NAV  0.054578   0.015293   3.569 0.000383 ***
## MOTIVATION   0.011745   0.002531   4.641 4.12e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4232 on 716 degrees of freedom
## Multiple R-squared:  0.05296,    Adjusted R-squared:  0.04635
## F-statistic: 8.008 on 5 and 716 DF,  p-value: 2.33e-07
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.448	0.075	5.965	0.000
DROPOUT_TOT	-0.049	0.013	-3.775	0.000
DROPOUT_INT	0.047	0.015	3.100	0.002
DROPOUT_LEA	0.047	0.016	3.034	0.003
DROPOUT_NAV	0.055	0.015	3.569	0.000
MOTIVATION	0.012	0.003	4.641	0.000

9 Block Stepwise Analysis (Target vs Profile + Course + Engagement - PRE)

9.1 Certificate Downloads

```
##
## Call:
## lm(formula = CERTIFICATE ~ GENDER + LANGUAGE + AGE + SECTOR +
##     DROPOUT_TOT + DROPOUT_INT + DROPOUT_NAV + MOTIVATION, data = General_C)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7843 -0.3504 -0.1972  0.4964  1.0036
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.071933   0.090984   0.791 0.429433
## GENDER      -0.173631   0.034147  -5.085 4.71e-07 ***
## LANGUAGE    -0.199914   0.057824  -3.457 0.000578 ***
## AGE          0.033790   0.007638   4.424 1.12e-05 ***
## SECTOR       0.007750   0.004573   1.695 0.090552 .
## DROPOUT_TOT -0.007826   0.004758  -1.645 0.100458
## DROPOUT_INT  0.021351   0.012091   1.766 0.077845 .
## DROPOUT_NAV  0.017364   0.010661   1.629 0.103807
```

```
## MOTIVATION   0.009928   0.002705   3.670 0.000260 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4519 on 713 degrees of freedom
## Multiple R-squared:  0.1075, Adjusted R-squared:  0.09747
## F-statistic: 10.73 on 8 and 713 DF,  p-value: 2.536e-14
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.072	0.091	0.791	0.429
GENDER	-0.174	0.034	-5.085	0.000
LANGUAGE	-0.200	0.058	-3.457	0.001
AGE	0.034	0.008	4.424	0.000
SECTOR	0.008	0.005	1.695	0.091
DROPOUT_TOT	-0.008	0.005	-1.645	0.100
DROPOUT_INT	0.021	0.012	1.766	0.078
DROPOUT_NAV	0.017	0.011	1.629	0.104
MOTIVATION	0.010	0.003	3.670	0.000

9.2   Completion Rate

```
##
## Call:
## lm(formula = CRATE ~ GENDER + LANGUAGE + AGE + CTUTORED + CCAT +
##     DROPOUT_TOT + DROPOUT_INT + DROPOUT_LEA + DROPOUT_NAV + MOTIVATION,
##     data = General_R)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9017 -0.3801  0.0102  0.3875  0.7882
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.327627   0.087425   3.748 0.000193 ***
## GENDER      -0.114899   0.032349  -3.552 0.000408 ***
## LANGUAGE    -0.121583   0.052992  -2.294 0.022061 *
## AGE          0.024805   0.006740   3.680 0.000250 ***
## CTUTORED     0.116756   0.040457   2.886 0.004021 **
## CCAT         0.094091   0.039909   2.358 0.018662 *
## DROPOUT_TOT -0.051726   0.012868  -4.020 6.45e-05 ***
## DROPOUT_INT  0.053442   0.015163   3.524 0.000452 ***
## DROPOUT_LEA  0.047758   0.015281   3.125 0.001848 **
## DROPOUT_NAV  0.059487   0.015039   3.955 8.40e-05 ***
## MOTIVATION   0.011440   0.002488   4.597 5.07e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4133 on 711 degrees of freedom
## Multiple R-squared:  0.1031, Adjusted R-squared:  0.09047
## F-statistic: 8.171 on 10 and 711 DF,  p-value: 1.365e-12
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.328	0.087	3.748	0.000
GENDER	-0.115	0.032	-3.552	0.000
LANGUAGE	-0.122	0.053	-2.294	0.022
AGE	0.025	0.007	3.680	0.000
CTUTORED	0.117	0.040	2.886	0.004
CCAT	0.094	0.040	2.358	0.019
DROPOUT_TOT	-0.052	0.013	-4.020	0.000
DROPOUT_INT	0.053	0.015	3.524	0.000
DROPOUT_LEA	0.048	0.015	3.125	0.002
DROPOUT_NAV	0.059	0.015	3.955	0.000
MOTIVATION	0.011	0.002	4.597	0.000