商管統計資料分析 期末報告

台灣之星 DCB 代收

第三組：

吳煜芬、姚宗志、彭盛皓

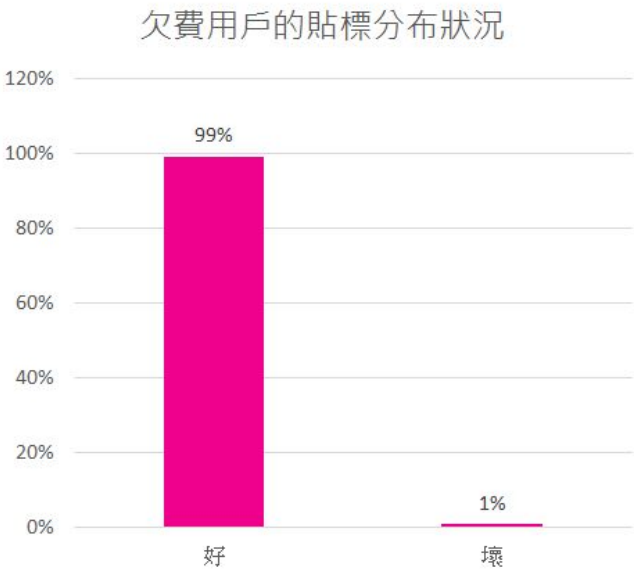
彭碩之、江庭萱、陳佳卉

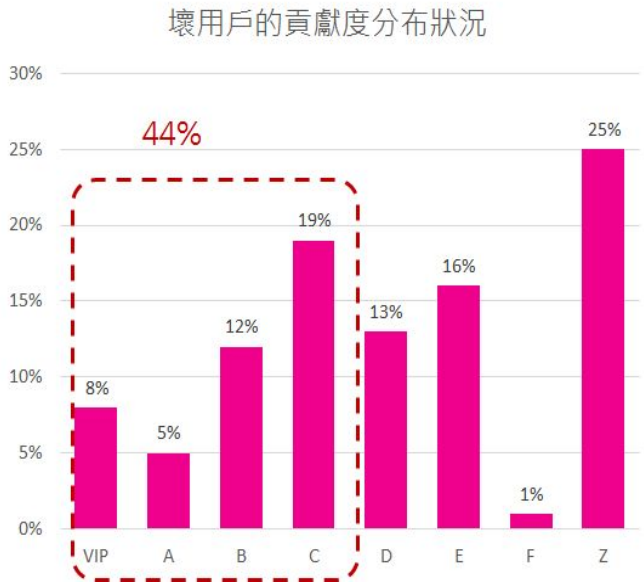
目錄

一、研究問題

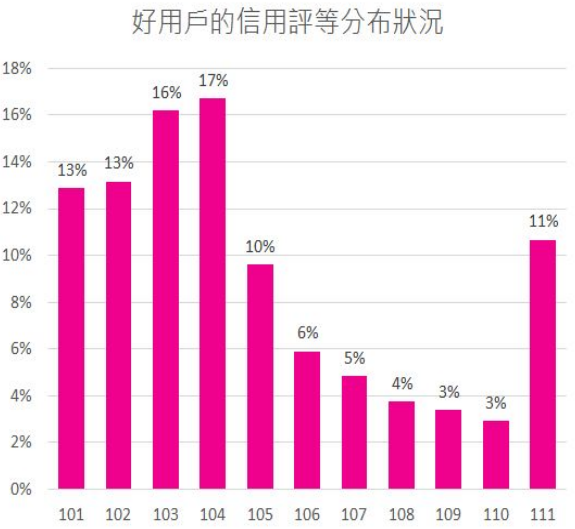
此次報告以DCB代收服務為主題，本組希望透過分析目前區分好壞用戶的貼標模型，並嘗試加入其他分群變數，建立以DCB為核心建立的用戶分群標準，以優化代收服務。

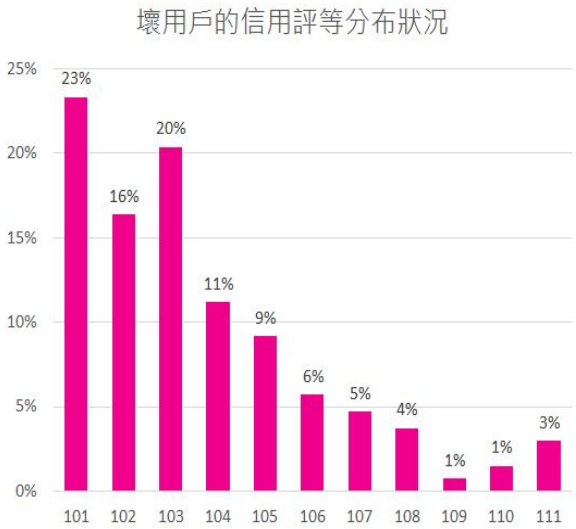
（一）EDA 研究





目前的好壞用乎分群定義為 - 出帳後2個月未繳款者設為壞用戶，本組深入分析好壞用戶其他項變數後，發現有44%的壞用戶為前段貢獻度用戶(VIP,A,B,C)，此外，在欠費用戶中(欠費定義為 - 近三個月每個月平均繳費金額 - 近3期平均出帳金額 < 0) 僅有1%的用戶被定義為壞用戶。顯現目前的好壞用戶分群並非針對DCB用戶，無法正確解釋用戶的繳費、貢獻度表現。





再分析好壞用戶的信用評等分布狀況，好用戶中約有75%分布在評等後段(101-106)，壞用戶中則約有85%，顯現好壞用戶在信用評等皆為較保守的分布。為研究目前信用評等的是否過度保守，由近三個月失敗平均購買金額佔信用等級比率變數分析，好用戶中約有25%的用戶曾經購買失敗，失敗金額約占信用等級比率的12%，平均失敗金額為87元。

進一步分析這25%的用戶，發現有48%為前段貢獻度用戶(VIP,A,B,C)，91%為近三個月整理DCB用戶消費前50%，顯現目前分群只有將用戶分成好、壞兩種較為簡單，目前的好用戶信用評級偏向保守。

二、研究目標

建立以DCB為核心建立的用戶分群標準，納入更多變向來衡量好壞用戶的分級，以降低欠繳比例；及藉由分析用戶行為提高DCB服務使用，來優化代收服務。

三、研究方法

（一）RFM 研究

依照最近一次購買天數 (R) 和購買頻率 (F)、消費金額 (M)，將顧客分為不同客群，並觀察其和性別、商品類別的關係。

（二）LM 和 GLM 分析

使用 R 中的線性模型(Linear Model) 和 廣義線性模型(General Linear Model)，以變數PAY（近三個月每個月平均繳費金額 - 近三期平均出帳金額）和其衍生的變數M（將PAY依0為區分點，區分好壞用戶）作為代表好壞用戶的反應變數，尋找相關的解釋變數。再利用這些相關的解釋變數，嘗試分辨好壞用戶。

四、研究過程與發現

（一）RFM 研究

依照最近一次購買天數 (R) 、購買頻率 (F)、消費金額 (M)，將顧客分為不同客群，並觀察其和性別、商品類別的關係。

1. 資料處理

首先將資料按照這三個指標進行處理：

R：使用各商品種類最近一次購買的累計天數，觀察資料分佈切分為四份

- 大於20天及NA值編碼為1
- 14-20天者編碼為2
- 7-13天者編碼為3
- 1-6天者編碼為4

編碼越大者為最近購買的用戶，將其定義為表現良好的用戶。

使用變數：

最近一次購買交友軟體的累計天數 G_LAST_BUY_DATE_APP_DAYS

最近一次購買遊戲的累計天數 G_LAST_BUY_GAME_APP_DAYS

最近一次購買直播的累計天數 G_LAST_BUY_LIVE_STREAM_DAYS

F：使用各商品種類的近三個月購買次數，觀察資料分佈切分為四份

- 小於3次者編碼為1
- 3-10次者編碼為2
- 11-20次者編碼為3
- 20次以上者編碼為4

編碼越大者為購買越多次的用戶，將其定義為表現良好的用戶。

使用變數：

用戶近三個月購買直播軟體次數 G_N3_BUY_LIVE_STREAM_CNT

用戶近三個月購買交友軟體次數 G_N3_BUY_DATE_APP_CNT

用戶近三個月購買遊戲軟體次數 G_N3_BUY_GAME_APP_CNT

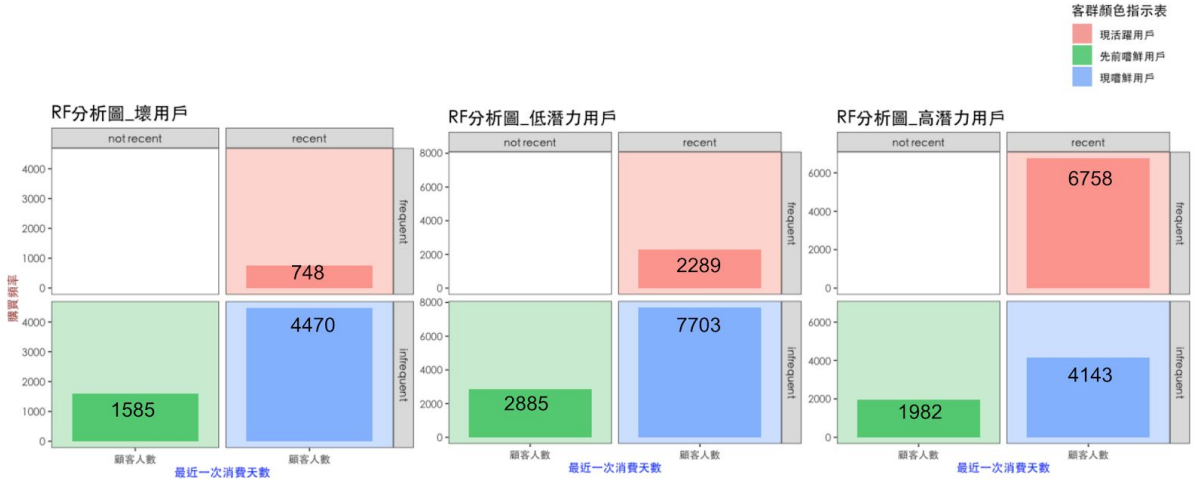
M：將 *近三個月每個月平均繳費金額 N3_PAID_MONEY_PER_MONTH* - *近3期平均出帳金額 L3MN_AVG_BILL_EXCL_PNLTY_AMT*，定義新的變數PAY

- PAY小於0者代表他繳得比出帳金額還少，定義為壞用戶並編碼為1
- 等於0者代表他沒有多繳也沒有少繳，定義為低潛力用戶並編碼為2
- 大於0者代表他繳得比出帳金額還多，定義為高潛力用戶並編碼為3

編碼越大者代表他繳得比出帳金額更多，將其定義為表現良好的用戶。

2. 各用戶群的RF分析

依據前述變數PAY定義的三種用戶群：高潛力用戶、低潛力用戶、壞用戶，分析各用戶群內R與F的人數分佈。



利用R和F進一步將用戶群內再分為4群：

- recent：兩週內有購買（R編碼3和4）、not recent：兩週內沒有購買（R編碼1和2）
- frequent：三個月內購買超過10次 （F編碼3和4）、 infrequent為購買小於10次（F編碼1和2）

再將這4群用戶定義為：現活躍用戶、先前嚐鮮用戶及現嚐鮮用戶

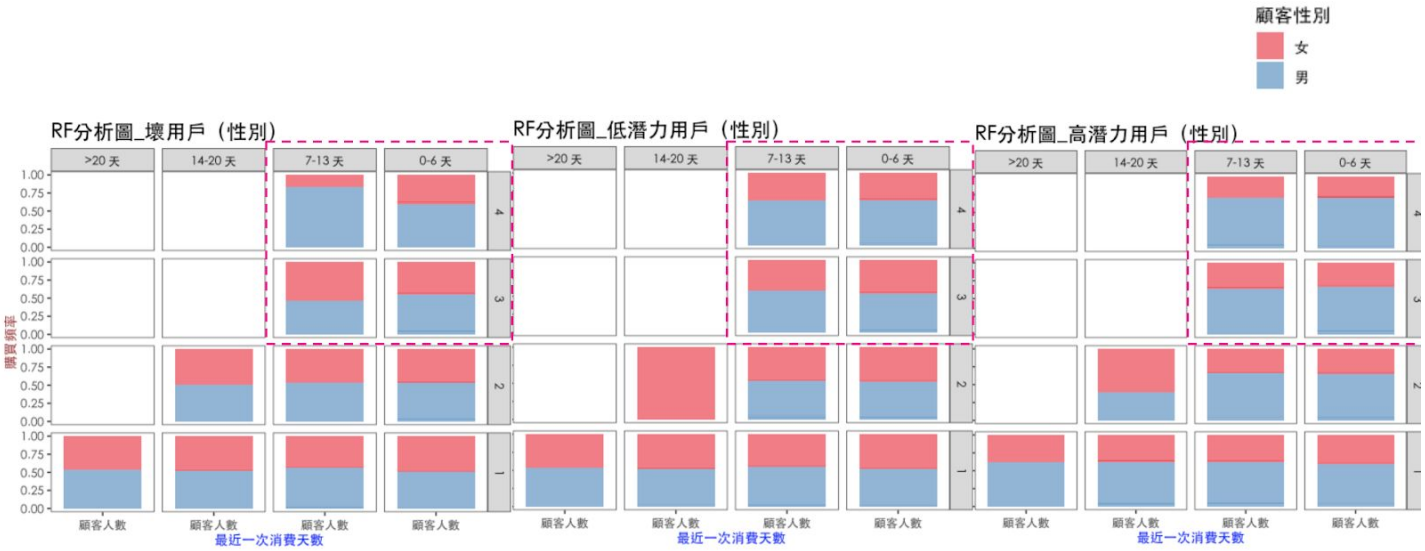
- 現活躍用戶：最近有購買、購買次數多，屬於表現最佳的用戶。
- 先前嚐鮮用戶：最近沒有購買、總次數多，之前常買但最近沒有買的用戶。
- 現嚐鮮用戶：最近有買、購買次數不多，之前沒有買但最近常買的用戶。

由圖表中得知，高潛力用戶中有52%為現活躍用戶。而低潛力和壞用戶的分佈類似，均以現嚐鮮用戶的比例最高，在低潛力用戶中佔60%，在壞用戶中則佔66%。

此外，根據各群均沒有not recent/frequent的用戶來看，推論消費次數多（三個月內大於10次）的用戶會維持購買習慣，不會超過兩週沒有購買。

3. 各用戶群的RF與性別分析

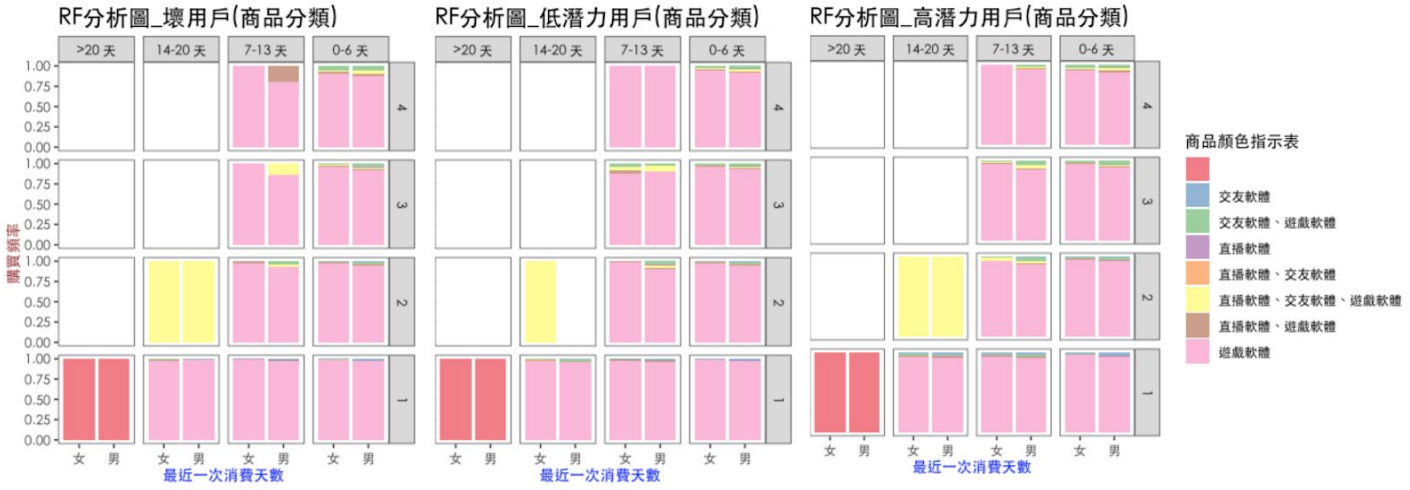
根據三種用戶群，按照R和F的1-4編碼分群，並呈現男女比例分佈狀況。



由圖表得知，不論是哪個用戶群，現活躍用戶皆以男性比例較高。而圖表中越往右上代表該群中R和F表現更好的用戶，也以男性的比例較高。

4. 各用戶群的RF與商品類別分析

根據三種用戶群，按照R和F的1-4編碼分群，並呈現購買商品類別的比例分佈。

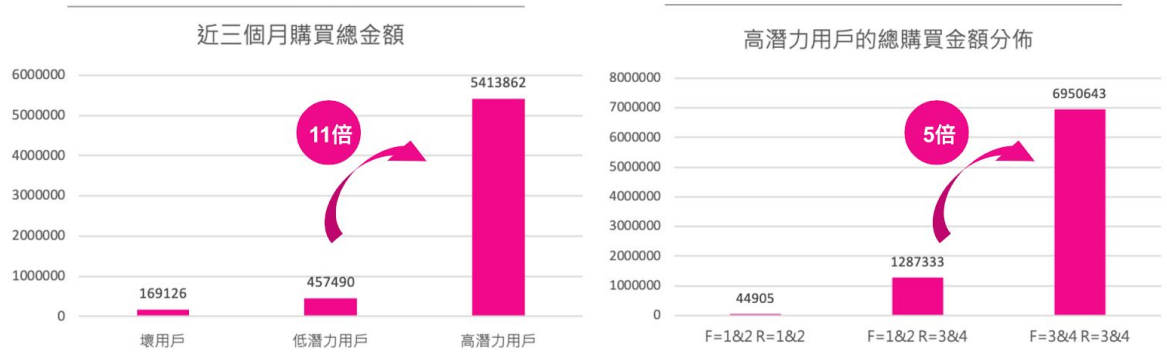


不論是哪個用戶群，多數用戶皆以購買遊戲軟體為主；而圖表中藍色區塊（只購買交友軟體）只出現在F為1和2時，可知只買交友軟體的用戶購買頻率少。

此外，若比較左右男女的分佈差異，發現現活躍用戶中男性購買遊戲以外商品的比例較女性高。

5. 結論

近期消費且頻率高的高潛力用戶能帶來龐大收益；高潛力的活躍用戶收益也大於嚐鮮用戶。



我們將F=4; R=4的各分群用戶的近三個月的購買總金額，發現高潛力用戶(M=3)的消費金額遠高於其他兩者分群；再深入探討高潛力用戶的購買金額，可發現現活躍用戶(F=3&4; R=3&4)的總金額也高於嚐鮮用戶(F=1&2)五倍以上；我們建議可針對高潛力用戶，提升 R 及 F 使嚐鮮用戶成為現活躍用戶。

(i) 針對F=1&2; R=3&4的高潛力用戶：提升用戶的消費頻率。

提供 DCB 服務優惠，例如購買可提供電話費優惠。

(ii) 針對F=1&2; R=1&2的高潛力用戶：提升用戶的消費頻率並積極提醒消費需求。

針對上次購買的商品類型進行簡訊推播，提醒消費；並提供DCB服務優惠。

（二）LM 和 GLM 分析

資料前處理：

- 類別化類別變數
- 空值處理：補0
- 刪除完全沒有變化的變數 (例如：員工優惠門號註記、欠費、更換資費方案等)

1. 初步建構線性模型

首先，我們考量到Google通路用戶會有 DCB的交易資訊，因此以兩種方式分別建構初步的線性模型：

a. **模型一**：針對全通路(即原資料集)做總概的屬性、帳款相關分析(不含DCB交易資訊)。

b. **模型二**：從原資料集中篩選DCB_CHANEL為Google的子資料集，針對Google通路用戶做屬性、帳款、DCB交易資訊的分析。

在模型一當中，使用PAY當應變數。

用戶屬性中納入解釋變數有CUR_AGE(目前年齡)、TNR(累積租期)、MAIN_MSISDN(主門號註記)、NP_OUT_GRP(新申裝或攜碼業者)等，其餘則因無特殊意義或納入後對模型影響甚微而剔除。

在繳款資訊部分，由於欠費、溢繳等問題有較高可能與應變數產生共線性，因此主要選取的變數為同樣是以三個月尺度來看的N3_PAID_CHANNEL_CATE(近三個月主要繳費通路)。

在帳單細項資訊方面，由於各項之間有較多重複，最後決定以近三個月平均每月的各項消費作為解釋變數，其中，由於國際簡訊費納入模型會不顯著，因此最後選取的變數有N3_VAS_FEE_PER_MON(近三個月平均每個月加值服務消費金額)、N3_MSG_FEE_PER_MON(近三個月平均每個月的簡訊費)。

在帳務資訊上，除了出帳金額與CNTRB_LVL(出帳時用戶貢獻度)外，我們試圖類別化出帳時經過隱碼的手機廠牌(HANDSET)，以評估其對應變數之影響；而月底資訊同樣考量到有多元共線性的可能，因此只選RATE_PLAN_RNG(月底專案實收總資費級距)作為解釋變數。

而在DCB額度資訊上，主要有進線與使用額度兩大類的資料，因此我們選擇各採取一個指標納入模型中，分別為：L3_AVG_MODI_CNT(近三個月平均每月進線調整次數)、L3_AVG_USED_QUOTA(近三個月每月平均使用額度)

在DCB交易資訊上，則納入了其近三個月的購買狀況，包含了N3_MONTH_PRCH_CNT_PER_MONTH(近三個月平均購買次數)、N3_MONTH_PRCH_AMOUNT_PER_MONTH(近三個月平均購買金額)

另外，我們也想了解購買程度在全體排名當中的位置與我們所設定的應變數之間的關係，因此採用了L3_AVG_AMT_NITILE_OVER_50_FLAG(是否為近三個月整體DCB用戶消費前50%)，以N為reference level，探討其與NA、Y之間的截距差。

模型一結果如下：

```
                                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                                       8.272e+01  5.043e+00  16.404  < 2e-16 ***
CUR_AGE(2) 20-29歲                                8.776e+00  1.528e+00   5.742 9.43e-09 ***
CUR_AGE(3) 30-39歲                                1.361e+01  1.598e+00   8.513  < 2e-16 ***
CUR_AGE(4) 40-49歲                                1.360e+01  1.595e+00   8.528  < 2e-16 ***
CUR_AGE(5) 50-59歲                                9.890e+00  1.784e+00   5.543 2.99e-08 ***
CUR_AGE(6) 大於60歲                               1.293e+01  2.308e+00   5.601 2.14e-08 ***
TNR                                               5.904e+00  2.396e-01  24.640  < 2e-16 ***
MAIN_MSISDNY                                     -1.187e+02  2.648e+00 -44.824  < 2e-16 ***
NP_OUT_GRP2                                      -1.563e+00  1.231e+00  -1.270 0.204112
NP_OUT_GRP3                                       8.020e+00  1.114e+00   7.198 6.18e-13 ***
NP_OUT_GRP4                                       5.495e+00  1.156e+00   4.753 2.01e-06 ***
NP_OUT_GRP5                                       7.167e+00  2.005e+00   3.575 0.000350 ***
N3_PAID_CHANNEL_CATE超商_全家                     2.677e+00  1.855e+00   1.443 0.149077
N3_PAID_CHANNEL_CATE超商_OK                       2.089e+00  3.217e+00   0.649 0.516085
N3_PAID_CHANNEL_CATE電通加盟                      8.392e+00  2.041e+00   4.112 3.93e-05 ***
N3_PAID_CHANNEL_CATE銀行                          3.177e+00  1.954e+00   1.626 0.103914
N3_PAID_CHANNEL_CATE郵局                         -2.164e+00  4.802e+00  -0.451 0.652177
N3_PAID_CHANNEL_CATE直營門市                      9.885e+00  1.902e+00   5.196 2.05e-07 ***
N3_PAID_CHANNEL_CATEACH                          -4.045e+00  5.349e+00  -0.756 0.449517
N3_PAID_CHANNEL_CATEOTHER                         1.910e+00  1.067e+01   0.179 0.857924
N3_PAID_CHANNEL_CATEOVERPAYMENT                  -1.982e+02  2.997e+00 -66.151  < 2e-16 ***
N3_VAS_FEE_PER_MON                               -7.184e-02  1.138e-02  -6.312 2.78e-10 ***
N3_MSG_FEE_PER_MON                               -1.021e-01  2.098e-02  -4.866 1.14e-06 ***
CNTRB_LVLB                                        4.536e+00  2.109e+00   2.151 0.031510 *
CNTRB_LVLC                                        2.312e+00  2.302e+00   1.004 0.315292
CNTRB_LVLD                                       -1.257e+01  2.534e+00  -4.959 7.12e-07 ***
CNTRB_LVLE                                       -2.624e+01  2.651e+00  -9.897  < 2e-16 ***
CNTRB_LVLF                                       -9.530e+00  3.764e+00  -2.532 0.011353 *
CNTRB_LVLVIP                                      3.076e+00  2.259e+00   1.362 0.173352
CNTRB_LVLZ                                       -2.917e+01  2.362e+00 -12.347  < 2e-16 ***
HANDSET2                                         -8.640e+00  1.772e+00  -4.875 1.09e-06 ***
HANDSET3                                         -6.738e+00  1.934e+00  -3.485 0.000493 ***
HANDSET4                                          2.255e+00  4.409e+00   0.511 0.609087
HANDSET5                                         -4.275e+00  5.820e+00  -0.735 0.462616
HANDSET6                                         -1.319e+01  1.623e+00  -8.125 4.56e-16 ***
HANDSET7                                          1.361e+01  2.225e+00   6.115 9.74e-10 ***
HANDSET8                                         -6.651e+00  1.528e+00  -4.353 1.35e-05 ***
HANDSET9                                         -4.433e+00  2.271e+00  -1.952 0.050965 .
HANDSET10                                        -9.163e+00  2.102e+00  -4.359 1.31e-05 ***
RATE_PLAN_RNG(2) L : 201~500                     -1.869e+01  2.441e+00  -7.656 1.95e-14 ***
RATE_PLAN_RNG(3) ML : 501~700                    -3.317e+01  2.732e+00 -12.142  < 2e-16 ***
RATE_PLAN_RNG(4) M : 701~900                     -5.884e+01  2.768e+00 -21.257  < 2e-16 ***
RATE_PLAN_RNG(5) MH : 901 以上                   -5.697e+01  3.033e+00 -18.781  < 2e-16 ***
L3_AVG_MODI_CNT                                   4.832e+00  6.695e-01   7.218 5.36e-13 ***
L3_AVG_USED_QUOTA                                 3.501e-01  3.436e-03 101.879  < 2e-16 ***
N3_MONTH_PRCH_CNT_PER_MONTH                      -1.922e-01  5.563e-02  -3.455 0.000551 ***
N3_MONTH_PRCH_AMOUNT_PER_MONTH                    2.365e-01  3.951e-03  59.858  < 2e-16 ***
L3_AVG_AMT_NITILE_OVER_50_FLAG                    1.757e+01  8.122e+00   2.164 0.030506 *
L3_AVG_AMT_NITILE_OVER_50_FLAGY                  -3.189e+01  9.933e-01 -32.104  < 2e-16 ***
N3_PEAK_TIME_REFUND_NTILE_50_AMT_PER_MONTH_FLAG   2.375e+00  1.311e+00   1.812 0.070006 .
N3_PEAK_TIME_REFUND_NTILE_50_AMT_PER_MONTH_FLAGY  5.022e-01  5.882e+00   0.085 0.931958
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 84.81 on 50064 degrees of freedom
Multiple R-squared:  0.7144,    Adjusted R-squared:  0.7142
```

模型一總結與詮釋及需要改進之處：

總結來看，此模型之 R square 與 Adjusted R square皆約0.714左右，由於應變數Y之分佈本身有負偏態之情形，導致QQ-plot也有一定程度之偏差。但是 Global usefulness test 和各變數的 Individual t test 都是顯著的。

在**年齡層**上，以20歲以下者作為reference level，各年齡層相對於我們所選定之指標皆有顯著差異；而**累積租期**越長，對指標PAY亦有正面影響；至於有**主門號註記**者反而使指標下降，推測可能與非主門號手機反而會是進行代收消費的客群有關。被隱碼的**新申裝與攜碼客戶**經類別化變數後，除了2以外，其餘皆與reference level 1有顯著差異。在**繳費通路**方面，以7-11為reference level，可發現電通加盟與直營門市繳費者相對而言有顯著差異，應該是需要積極經營的高潛力客戶。

此外，平均每個月的**加值服務消費金額**與**簡訊消費金額**則會使指標PAY下降，可能與消費者的預算限制有所關聯。至於**出帳時用戶貢獻度**以A作為reference level時與其他level間有顯著性不等的差異，因此可考慮改變參考基準後再次衡量此變數對於應變數的影響。另外，被隱碼的**出帳時手機廠牌**對指標PAY亦有顯著性不等之差異，建議可進一步針對品牌了解之。在**月底專案實收總資費級距**的方面，級距越高者會使指標PAY越低，亦是可進一步分析之議題。

在額度資訊上，無論是**平均每月進線調整次數**與**平均每月使用額度**對指標PAY而言皆有正面影響，至於購買次數對指標PAY有負面影響、購買金額則又對PAY有正面影響，則可能需再進一步調整模型或分析。在「是否為近三個月整體DCB用戶消費前50%」這個議題上，以N為reference level時，其與NA或Y之間的差距顯著，單就這個變數相比之下，資料為NA或N者或許是更具潛力之用戶。

模型二也使用PAY當應變數，在模型一所使用的變數基礎上，根據變數顯著性選取一些變數，並加上DCB交易資訊變數，研究PAY和Google Store上面消費的關係。

其中增加的DCB交易資訊變數為：

近三個月購買後50%（前5%）熱門GOOGLE商品次數

G_N3_BUY_BUTTOM50_CNT_CNT

G_N3_BUY_TOP5_CNT_CNT

近三個月是否購買後50%（前5%）熱門GOOGLE商品

G_N3_BUY_BUTTOM50_CNT_OR_NOT

G_N3_BUY_TOP5_CNT_OR_NOT

近三個月是否購買後50%（前5%）GOOGLE商品平均金額

G_N3_BUY_BUTTOM50_CNT_AVG_AMT

G_N3_BUY_TOP5_CNT_AVG_AMT

近三個月是否購買後50%（前5%）購買人次GOOGLE商品

G_N3_BUY_BUTTOM50_DCON_OR_NOT

G_N3_BUY_TOP5_DCON_OR_NOT

近三個月購買後50%（前5%）平均金額人次GOOGLE商品

G_N3_BUY_BUTTOM50_AVG_PRICE_OR_NOT

G_N3_BUY_TOP5_AVG_PRICE_OR_NOT

上面變數主要研究後50%（前5%）的GOOGLE商品對PAY的影響

用戶近三個月是否購買交友軟體 G_N3_BUY_DATE_APP_OR_NOT

用戶近三個月是否購買遊戲軟體 G_N3_BUY_GAME_APP_OR_NOT

近三個月一般時刻平均每月成功消費金額

N3_NORM_TIME_CNT_PER_MONTH

模型二結果如下：

```
Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                         4.954e+01  8.568e+00   5.782 7.46e-09 ***
CUR_AGE(2) 20-29歲                  8.454e+00  2.163e+00   3.908 9.31e-05 ***
CUR_AGE(3) 30-39歲                  7.023e+00  2.185e+00   3.215 0.001307 **
CUR_AGE(4) 40-49歲                  6.922e+00  2.169e+00   3.191 0.001421 **
CUR_AGE(5) 50-59歲                  4.577e+00  2.390e+00   1.915 0.055490 .
CUR_AGE(6) 大於60歲                 1.001e+01  2.998e+00   3.339 0.000843 ***
N3_PAID_CHANNEL_CATEOTHER           1.226e+01  1.602e+01   0.766 0.443898
N3_PAID_CHANNEL_CATEOVERPAYMENT    -1.918e+02  7.855e+00 -24.422  < 2e-16 ***
N3_PAID_CHANNEL_CATE直營門市        2.483e+01  7.071e+00   3.512 0.000446 ***
N3_PAID_CHANNEL_CATE超商_OK         1.427e+01  7.775e+00   1.836 0.066389 .
N3_PAID_CHANNEL_CATE超商_全家       1.574e+01  7.047e+00   2.234 0.025510 *
N3_PAID_CHANNEL_CATE超商_萊爾富     1.439e+01  7.367e+00   1.953 0.050839 .
N3_PAID_CHANNEL_CATE郵局            5.508e+00  9.102e+00   0.605 0.545056
N3_PAID_CHANNEL_CATE銀行            1.149e+01  7.116e+00   1.615 0.106405
N3_PAID_CHANNEL_CATE電通加盟        2.393e+01  7.127e+00   3.357 0.000788 ***
MAIN_MSISDNY                       -1.184e+02  3.455e+00 -34.285  < 2e-16 ***
N3_OVER_PAY_FLAGY                   1.792e+01  1.755e+00  10.209  < 2e-16 ***
N3_BUY_AMT_PER_QUOTA               -6.405e+00  1.701e-01 -37.654  < 2e-16 ***
N3_MONTH_PRCH_CNT_PER_MONTH        -1.704e+00  2.351e-01  -7.247 4.35e-13 ***
N3_MONTH_PRCH_AMOUNT_PER_MONTH      6.341e-01  5.395e-03 117.540  < 2e-16 ***
N3_MSG_FEE_PER_MON                 -8.396e-02  2.602e-02  -3.227 0.001254 **
CNTRB_LVLB                          1.927e-01  2.612e+00   0.074 0.941193
CNTRB_LVLC                          5.043e-01  2.516e+00   0.200 0.841160
CNTRB_LVLD                          8.078e-01  2.620e+00   0.308 0.757834
CNTRB_LVLE                         -4.290e+00  2.547e+00  -1.685 0.092082 .
CNTRB_LVLF                          1.993e+01  3.426e+00   5.817 6.03e-09 ***
CNTRB_LVLVIP                        5.322e+00  3.000e+00   1.774 0.076040 .
CNTRB_LVLZ                         -2.391e+01  2.732e+00  -8.750  < 2e-16 ***
TNR                                 4.435e+00  2.881e-01  15.394  < 2e-16 ***

L3_AVG_MODI_CNT                     7.914e+00  8.561e-01   9.244  < 2e-16 ***
G_N3_BUY_BUTTOM50_CNT_CNT           8.129e-01  2.157e-01   3.768 0.000165 ***
G_N3_BUY_TOP5_CNT_CNT               3.888e-01  3.065e-02  12.682  < 2e-16 ***
G_N3_BUY_BUTTOM50_CNT_OR_NOTY       1.737e+00  3.656e+00   0.475 0.634708
G_N3_BUY_TOP5_CNT_OR_NOTY           3.455e+01  3.763e+00   9.181  < 2e-16 ***
G_N3_BUY_GAME_APP_OR_NOTY          -8.242e+00  1.601e+00  -5.149 2.64e-07 ***
G_N3_BUY_DATE_APP_OR_NOTY          -1.137e+01  3.060e+00  -3.715 0.000203 ***
G_N3_BUY_BUTTOM50_CNT_AVG_AMT      -8.503e-02  1.312e-02  -6.479 9.37e-11 ***
G_N3_BUY_TOP5_CNT_AVG_AMT          -2.969e-01  1.427e-02 -20.813  < 2e-16 ***
G_N3_BUY_BUTTOM50_DCON_OR_NOTY      4.101e+00  3.049e+00   1.345 0.178623
G_N3_BUY_BUTTOM50_AVG_PRICE_OR_NOTY 3.869e+00  1.711e+00   2.260 0.023802 *
G_N3_BUY_TOP5_DCON_OR_NOTY         -1.429e+01  3.268e+00  -4.372 1.23e-05 ***
G_N3_BUY_TOP5_AVG_PRICE_OR_NOTY     1.111e+01  2.930e+00   3.794 0.000149 ***
N3_NORM_TIME_CNT_PER_MONTH          1.524e+00  2.323e+00   0.656 0.511698
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 89.96 on 31168 degrees of freedom
  (1352 observations deleted due to missingness)
Multiple R-squared:  0.7064,    Adjusted R-squared:  0.706
F-statistic:  1786 on 42 and 31168 DF,  p-value: < 2.2e-16
```

模型二總結與詮釋及需要改進之處：

總結來看，此模型之 R square 與 Adjusted R square皆約 0.706 左右，由於應變數Y之分佈本身有負偏態之情形，導致QQ-plot也有一定程度之偏差。但是 Global usefulness test 和各變數的 Individual t test 都是顯著的。

1. 用戶的付款表現優異與否和用戶DCB的使用情況是有關聯的；新指標與大多的變數呈現顯著性，且變數係數與新指標的正/負關係皆為合理的，是能被參考的。

   (1) 年齡區間：**20-29歲**的用戶PAY表現較其他區間好

   (2) 付費通路：在**直營門市付費**的用戶PAY表現較其他通路突出，可以積極經營該族群使用DCB服務

2. 比較特別的發現為：

   若用戶購買排名後50%商品，並達到平均金額，他的付款表現是有顯著優異的；另外，透過DCB購買排名前5或後50%商品的次數越多，付款表現也是正向的，顯示**培養使用DCB習慣是重要的**，可以透過後續行銷策略引導消費者，增加收益。

2. 從 LM 到 GLM

雖然上述所建構的兩個模型就統計上而言可以對應變數PAY 作出合理解釋，但實際上並不便於分析，由於我們後來另外以 RFM模型的概念進行分析，因此決定修改新產生的衍生變數 M，將原先 M 編碼為 1 修改為 0，涵蓋的即是「壞用戶」（原先指標 PAY < 0 者）；並將原先 M 編碼為 2 與 3 修改為 1，涵蓋的即是「高潛力用戶」與「低潛力用戶」（原先指標 PAY≥0 者），並改為建構GLM模型。

先用模型一的解釋變數建構GLM模型，再利用Anova 表找出哪些變數對於 Deviance 的降低有較大貢獻；另外，與 DCB 相關的解釋變數也加進GLM模型中。

	Df	Deviance	Resid. Df	Resid. Dev
NULL			32562	33378
FLAG	1	9.6	32561	33369
CUR_AGE	5	259.1	32556	33110
TNR	1	2919.0	32555	30191
GNDR	1	91.2	32554	30099
MAIN_MSISDN	1	135.0	32553	29964
NP_OUT_GRP	1	23.4	32552	29941
N3_PAID_CHANNEL_CATE	9	992.4	32543	28949
N3_PAID_CNT_PER_MONTH	1	186.4	32542	28762
N3_OVER_PAY_FLAG	1	24.2	32541	28738
N3_INTERNATION_CALL_FLAG	1	1.6	32540	28736
N3_VAS_FEE_PER_MON	1	2.7	32539	28734
N3_MSG_FEE_PER_MON	1	0.4	32538	28733
N3_VAS_APPLY_FLAG	1	5.2	32537	28728
N3_INTERNATIONAL_MSG_FLAG	1	34.7	32536	28693
USE_STAT_NM	3	3.5	32533	28690
CNTRB_LVL	7	1113.4	32526	27577
HANDSET	1	0.2	32525	27576
PROJ_USE_HANDSET	1	513.5	32524	27063
ACTV_CLASS_NM	1	357.5	32523	26705
FEW_CALL_INDIC	1	0.0	32522	26705
NO_CALL_INDIC	1	0.1	32521	26705
BILL_AMT	1	4313.7	32520	22392
BILL_EXCL_PNLTY_AMT	1	1142.5	32519	21249
BILL_PRE_CHRG_AMT	1	1761.1	32518	19488
RATE_PLAN_RNG	4	118.9	32514	19369
TOTAL_MONTHLY_RATE	1	17.5	32513	19352
CURRENT_LEVEL	1	303.0	32512	19049
L3_PSUSPEND_FLAG	1	225.2	32511	18823
L3_SUSPEND_FLAG	1	2.1	32510	18821
AUTO_TRANS_FLAG	1	39.2	32509	18782
L3_AVG_LEVEL_DOWN_CNT	1	31.4	32508	18751
L3_MODI_MAX_LEVEL	1	18.7	32507	18732
L3_AVG_MODI_CNT	1	3.0	32506	18729
L3_MODI_FLAG	1	0.2	32505	18729
L3_AVG_LEVEL_UP_CNT	1	168.1	32504	18561
L3_MAX_LEVEL	1	14.7	32503	18546
L3_MAX_LEVEL_MONTH_DIFF	1	0.5	32502	18545
L3_AVG_USED_QUOTA	1	1090.5	32501	17455
L3_SERV_FLAG	1	12.1	32500	17443
N3_MONTH_PRCH_CNT_PER_MONTH	1	38.6	32499	17404
N3_MONTH_PRCH_AMOUNT_PER_MONTH	1	43.0	32498	17361
LAST_MONTH_PRCH_AMOUNT	1	0.0	32497	17361
N3_BUY_AMT_PER_QUOTA	1	1.3	32496	17360
LATEST_FAIL_PRICE_PER_QUOTA	1	5.6	32495	17354
L3_AVG_AMT_NITILE_OVER_50_FLAG	2	16.5	32493	17338
N3_FAIL_NORM_TIME_CNT_PER_MONTH	1	6.3	32492	17332
N3_FAIL_NORM_TIME_AMOUNT_PER_MONTH	1	0.0	32491	17332
N3_FAIL_PEAK_TIME_CNT_PER_MONTH	1	2.8	32490	17329
N3_FAIL_PEAK_TIME_AMOUNT_PER_MONTH	1	0.0	32489	17329
N3_PEAK_TIME_AMOUNT_PER_MONTH	1	0.8	32488	17328
N3_PEAK_TIME_NTILE_10_AMT_PER_MONTH_FLAG	1	0.1	32487	17328
N3_NORM_TIME_CNT_PER_MONTH	1	1.9	32486	17326
G_N3_BUY_BUTTOM50_CNT_CNT	1	0.0	32485	17326
G_N3_BUY_TOP5_CNT_CNT	1	14.5	32484	17312
G_N3_BUY_BUTTOM50_CNT_OR_NOT	1	5.1	32483	17306
G_N3_BUY_TOP5_CNT_OR_NOT	1	7.0	32482	17299
G_N3_BUY_GAME_APP_OR_NOT	1	0.2	32481	17299
G_N3_BUY_DATE_APP_OR_NOT	1	0.1	32480	17299
G_N3_BUY_BUTTOM50_CNT_AVG_AMT	1	1.8	32479	17297
G_N3_BUY_TOP5_CNT_AVG_AMT	1	3.1	32478	17294
G_N3_BUY_BUTTOM50_DCON_OR_NOT	1	1.2	32477	17293
G_N3_BUY_BUTTOM50_AVG_PRICE_OR_NOT	1	3.1	32476	17290
G_N3_BUY_TOP5_DCON_OR_NOT	1	0.3	32475	17290
G_N3_BUY_TOP5_AVG_PRICE_OR_NOT	1	0.5	32474	17289

最後從模型一選出：

TNR (累積租期)

N3_PAID_CHANNEL_CATE (近三個月主要繳費通路)

CNTRB_LVL (出帳時用戶貢獻度)

PROJ_USE_HANDSET (出帳時專案搭配手機廠牌)

BILL_AMT (當期出帳金額)

BILL_EXCL_PNLTY_AMT (當期出帳金額(不含違約金及儲值卡折抵))、

BILL_PRE_CHRG_AMT (當期出帳金額(含違約金))

CURRENT_LEVEL (信用等級)

L3_AVG_USED_QUOTA (近三個月每月平均使用額度)

MAIN_MSISDN (主門號註記)

以及從模型二中選出DCB交易資訊相關變數

GLM 模型結果如下：

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.1593   0.0022   0.1723   0.4790   4.6031

Coefficients: (7 not defined because of singularities)
                                    Estimate Std. Error z value Pr(>|z|)
(Intercept)                       -1.971e+00  1.079e+00  -1.827 0.067736 .
TNR                                2.242e-01  1.640e-02  13.675  < 2e-16 ***
N3_PAID_CHANNEL_CATEOTHER          3.741e-05  5.184e-01   0.000 0.999942
N3_PAID_CHANNEL_CATEOVERPAYMENT   -6.177e-01  2.767e-01  -2.232 0.025597 *
N3_PAID_CHANNEL_CATE直營門市       9.440e-01  2.270e-01   4.159 3.20e-05 ***
N3_PAID_CHANNEL_CATE超商_OK        6.616e-01  2.570e-01   2.575 0.010038 *
N3_PAID_CHANNEL_CATE超商_全家      6.752e-01  2.255e-01   2.994 0.002749 **
N3_PAID_CHANNEL_CATE超商_萊爾富    7.612e-01  2.381e-01   3.196 0.001392 **
N3_PAID_CHANNEL_CATE郵局           3.673e-01  2.943e-01   1.248 0.212015
N3_PAID_CHANNEL_CATE電通加盟       8.615e-01  2.300e-01   3.746 0.000180 ***
N3_PAID_CHANNEL_CATE銀行           5.805e-01  2.276e-01   2.550 0.010759 *
CNTRB_LVLB                        -9.074e-02  1.263e-01  -0.718 0.472489
CNTRB_LVLC                        -3.190e-01  1.227e-01  -2.600 0.009318 **
CNTRB_LVLD                        -7.297e-01  1.288e-01  -5.664 1.48e-08 ***
CNTRB_LVLE                        -1.361e+00  1.351e-01 -10.079  < 2e-16 ***
CNTRB_LVLF                        -1.181e+00  1.662e-01  -7.106 1.19e-12 ***
CNTRB_LVLVIP                       2.913e-01  1.578e-01   1.846 0.064901 .
CNTRB_LVLZ                        -7.967e-01  1.310e-01  -6.084 1.17e-09 ***
PROJ_USE_HANDSET2                  7.093e-01  1.393e-01   5.091 3.57e-07 ***
PROJ_USE_HANDSET3                  6.747e-01  1.531e-01   4.408 1.04e-05 ***
PROJ_USE_HANDSET4                  7.478e-01  1.397e-01   5.355 8.56e-08 ***
PROJ_USE_HANDSET5                  4.896e-01  2.236e-01   2.190 0.028542 *
PROJ_USE_HANDSET6                  6.400e-01  1.349e-01   4.743 2.10e-06 ***
PROJ_USE_HANDSET7                  5.562e-01  1.553e-01   3.580 0.000343 ***
```

```
PROJ_USE_HANDSET8                   9.389e-01  1.477e-01   6.355 2.09e-10 ***
PROJ_USE_HANDSET9                   9.386e-01  1.396e-01   6.722 1.79e-11 ***
BILL_AMT                            8.201e-03  4.960e-04  16.533  < 2e-16 ***
BILL_EXCL_PNLTY_AMT                 4.445e-02  1.438e-03  30.913  < 2e-16 ***
BILL_PRE_CHRG_AMT                  -5.897e-02  1.319e-03 -44.723  < 2e-16 ***
CURRENT_LEVEL                       5.075e-02  9.572e-03   5.301 1.15e-07 ***
L3_AVG_USED_QUOTA                   8.115e-03  2.860e-04  28.378  < 2e-16 ***
MAIN_MSISDNY                       -2.387e+00  1.895e-01 -12.599  < 2e-16 ***
G_N3_BUY_BUTTOM50_CNT_CNT           6.276e-02  2.752e-02   2.281 0.022576 *
G_N3_BUY_TOP5_CNT_CNT               2.326e-02  3.119e-03   7.455 8.97e-14 ***
G_N3_BUY_BUTTOM50_CNT_OR_NOT       -4.577e-01  4.389e-01  -1.043 0.297021
G_N3_BUY_BUTTOM50_CNT_OR_NOTY      -4.420e-01  1.808e-01  -2.445 0.014482 *
G_N3_BUY_TOP5_CNT_OR_NOT                   NA         NA      NA       NA
G_N3_BUY_TOP5_CNT_OR_NOTY          -3.843e-01  1.347e-01  -2.852 0.004343 **
G_N3_BUY_GAME_APP_OR_NOT                   NA         NA      NA       NA
G_N3_BUY_GAME_APP_OR_NOTY           4.543e-02  5.351e-02   0.849 0.395884
G_N3_BUY_DATE_APP_OR_NOT                   NA         NA      NA       NA
G_N3_BUY_DATE_APP_OR_NOTY           7.594e-02  1.454e-01   0.522 0.601562
G_N3_BUY_BUTTOM50_CNT_AVG_AMT       2.453e-04  9.001e-04   0.273 0.785202
G_N3_BUY_TOP5_CNT_AVG_AMT           2.312e-03  5.562e-04   4.157 3.22e-05 ***
G_N3_BUY_BUTTOM50_DCON_OR_NOT              NA         NA      NA       NA
G_N3_BUY_BUTTOM50_DCON_OR_NOTY      1.732e-01  1.379e-01   1.256 0.209068
G_N3_BUY_BUTTOM50_AVG_PRICE_OR_NOT         NA         NA      NA       NA
G_N3_BUY_BUTTOM50_AVG_PRICE_OR_NOTY 5.852e-02  6.177e-02   0.947 0.343460
G_N3_BUY_TOP5_DCON_OR_NOT                  NA         NA      NA       NA
G_N3_BUY_TOP5_DCON_OR_NOTY         -8.717e-02  1.203e-01  -0.725 0.468645
G_N3_BUY_TOP5_AVG_PRICE_OR_NOT             NA         NA      NA       NA
G_N3_BUY_TOP5_AVG_PRICE_OR_NOTY     7.209e-01  3.072e-01   2.346 0.018959 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 33378  on 32562  degrees of freedom
Residual deviance: 18032  on 32518  degrees of freedom
AIC: 18122

Number of Fisher Scoring iterations: 8
```

此模型中 deviance 從 33378 降至 18032。

3. 考慮DCB變數及多元共線性問題

DCB變數問題

我們可以從上述加入DCB變數的模型呈現結果中發現，加入DCB變數對於Deviance的影響很小，其中一個可能是DCB變數的資料量太少所致，經過考量後，我們決定GLM模型將不採納DCB相關變數。

多元共線性問題

在目前的模型下，我們最後利用GVIF（廣義變異數膨脹因子）的計算作GLM下多元共線性之檢查，結果如下：

```
                        GVIF Df GVIF^(1/(2*Df))
TNR                 1.568934  1        1.252571
N3_PAID_CHANNEL_CATE 1.073808  9        1.003964
CNTRB_LVL           5.759039  7        1.133211
PROJ_USE_HANDSET    2.516337  8        1.059371
BILL_AMT            5.137492  1        2.266604
BILL_EXCL_PNLTY_AMT 26.026821  1        5.101649
BILL_PRE_CHRG_AMT   20.500021  1        4.527695
CURRENT_LEVEL       1.244989  1        1.115791
L3_AVG_USED_QUOTA   1.971995  1        1.404277
MAIN_MSISDN         1.017605  1        1.008764
```

一般而言， VIF的閾值(threshold)訂為10，而上圖另外呈現的GVIF^(1/(2*df))閾值則會是sqrt(10)=3.16，從上圖可以發現，有部分變數的GVIF與GVIF^(1/(2*df))皆大於閾值，需要適時捨去，此GLM模型中，當期出帳金額有三個非常相關的變數，差別在於有無涵蓋違約金，而將另外兩個當期出帳金額捨去後，只留下一個後，再以剩下的變數計算GVIF後，最後模型應無多元共線性問題。

最終模型如下：

```
Call:
glm(formula = M ~ TNR + N3_PAID_CHANNEL_CATE + CNTRB_LVL + PROJ_USE_HANDSET +
    BILL_AMT + CURRENT_LEVEL + L3_AVG_USED_QUOTA + MAIN_MSISDN,
    family = binomial, data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.7290   0.0121   0.2447   0.5728   3.1292

Coefficients:
                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                     -1.068e+01  9.594e-01 -11.136  < 2e-16 ***
TNR                              3.622e-01  1.667e-02  21.726  < 2e-16 ***
N3_PAID_CHANNEL_CATEOTHER        1.756e-02  4.584e-01   0.038 0.969443
N3_PAID_CHANNEL_CATEOVERPAYMENT -9.079e-01  2.438e-01  -3.723 0.000197 ***
N3_PAID_CHANNEL_CATE直營門市     7.749e-01  2.103e-01   3.684 0.000229 ***
N3_PAID_CHANNEL_CATE超商_OK      5.644e-01  2.374e-01   2.378 0.017430 *
N3_PAID_CHANNEL_CATE超商_全家    5.905e-01  2.091e-01   2.824 0.004742 **
N3_PAID_CHANNEL_CATE超商_萊爾富  6.500e-01  2.203e-01   2.950 0.003176 **
N3_PAID_CHANNEL_CATE郵局         3.718e-01  2.742e-01   1.356 0.175087
N3_PAID_CHANNEL_CATE電通加盟     6.947e-01  2.129e-01   3.264 0.001099 **
N3_PAID_CHANNEL_CATE銀行         5.086e-01  2.110e-01   2.410 0.015943 *
CNTRB_LVLB                       2.245e-01  1.170e-01   1.920 0.054912 .
CNTRB_LVLC                       2.325e-01  1.119e-01   2.077 0.037807 *
CNTRB_LVLD                      -8.469e-02  1.152e-01  -0.735 0.462357
CNTRB_LVLE                      -6.436e-01  1.153e-01  -5.584 2.35e-08 ***
CNTRB_LVLF                       2.089e-01  1.392e-01   1.501 0.133378
CNTRB_LVLVIP                    -3.140e-01  1.460e-01  -2.151 0.031484 *
CNTRB_LVLZ                      -5.249e-01  1.154e-01  -4.548 5.41e-06 ***
PROJ_USE_HANDSET2                1.471e+00  1.176e-01  12.505  < 2e-16 ***
PROJ_USE_HANDSET3                1.628e+00  1.327e-01  12.261  < 2e-16 ***
PROJ_USE_HANDSET4                1.208e+00  1.166e-01  10.369  < 2e-16 ***
PROJ_USE_HANDSET5                1.225e+00  1.974e-01   6.205 5.48e-10 ***
PROJ_USE_HANDSET6                1.401e+00  1.137e-01  12.323  < 2e-16 ***
PROJ_USE_HANDSET7                1.368e+00  1.306e-01  10.471  < 2e-16 ***
PROJ_USE_HANDSET8                1.626e+00  1.251e-01  13.000  < 2e-16 ***
PROJ_USE_HANDSET9                2.745e+00  1.171e-01  23.435  < 2e-16 ***

BILL_AMT                         9.342e-03  2.875e-04  32.493  < 2e-16 ***
CURRENT_LEVEL                    7.462e-02  8.691e-03   8.586  < 2e-16 ***
L3_AVG_USED_QUOTA                4.180e-03  1.800e-04  23.224  < 2e-16 ***
MAIN_MSISDNY                    -2.028e+00  1.591e-01 -12.744  < 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: 33378  on 32562  degrees of freedom
Residual deviance: 21808  on 32533  degrees of freedom
AIC: 21868

Number of Fisher Scoring iterations: 7
```

最終模型中對於降低 deviance 的能力變弱是因為捨棄了造成多重共線性的變數(BILL_EXCL_PNLTY_AMT、BILL_PRE_CHRG_AMT)。

最後的GLM模型主要呈現累積租期、代收管道、出帳時用戶貢獻度、出帳時專案搭配手機品牌、當期出帳金額、信用評等、近三個月每月平均使用額度、主門號註記對於二元應變數M的影響，亦即預測我們所認定之「好用戶」、「壞用戶」的機率，加以貼標並改進代收業務的客群分析與效率，同時也可針對GLM模型中可能對指標M產生正面效果的變數作行銷策略的改進，試圖挖掘更多高潛力用戶。

4. 最終GLM 模型解釋

(1) 連續變數如 BILL_AMT (當期出帳金額)，舉例而言，出帳金額200者為好用戶的Odds會是出帳金額為100者的2.55倍 (exp(0.009342*(200-100)=2.55)。

(2) 類別變數如MAIN_MSISDN（主門號註記），以非主門號註記者為參考基準，主門號註記者為好用戶的Odds會是非主門號註記者的0.13倍 (exp(-2.028)=0.13)。

整體而言，當模型上的解釋變數係數為正值時，對於好用戶指標有正面影響；解釋變數係數為負值時則反之。最後，若我們匯入一筆真實資料至模型內時，所計算出的logit值再進行轉換即可得到二元變數M為1的機率。如此，能再進一步設定一個機率值為歸類二元變數的基準，舉例而言，若設定的機率值為0.5，當計算出的預測機率值大於0.5時，即被我們歸類為「好用戶」；計算出的預測機率值小於0.5時，即被我們歸類為「壞用戶」，我們也就是根據這個方法優化代收的分群基準。

五、結論

RFM：

針對高潛力用戶群，提升 R 及 F 使嚐鮮用戶成為現活躍用戶

1. 提供 DCB 服務優惠，例如購買可提供電話費優惠
2. 針對上次購買的商品類型進行簡訊推播，提醒消費

GLM：

優化具潛力用戶指標，以模型解釋變數結果進行分析，拓展行銷策略

1. 加強非主門號註記用戶之經營，挖掘更多具潛力用戶
2. 針對直營門市、電通加盟作為代收管道之目標目標客群設計行銷方案