



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

問題

現有的Flag分群非針對代收服務，無法精確反映用戶表現

目標

建立以DCB為核心建立的用戶分群標準，並藉由分析用戶行為提高DCB服務使用，來優化代收服務

研究

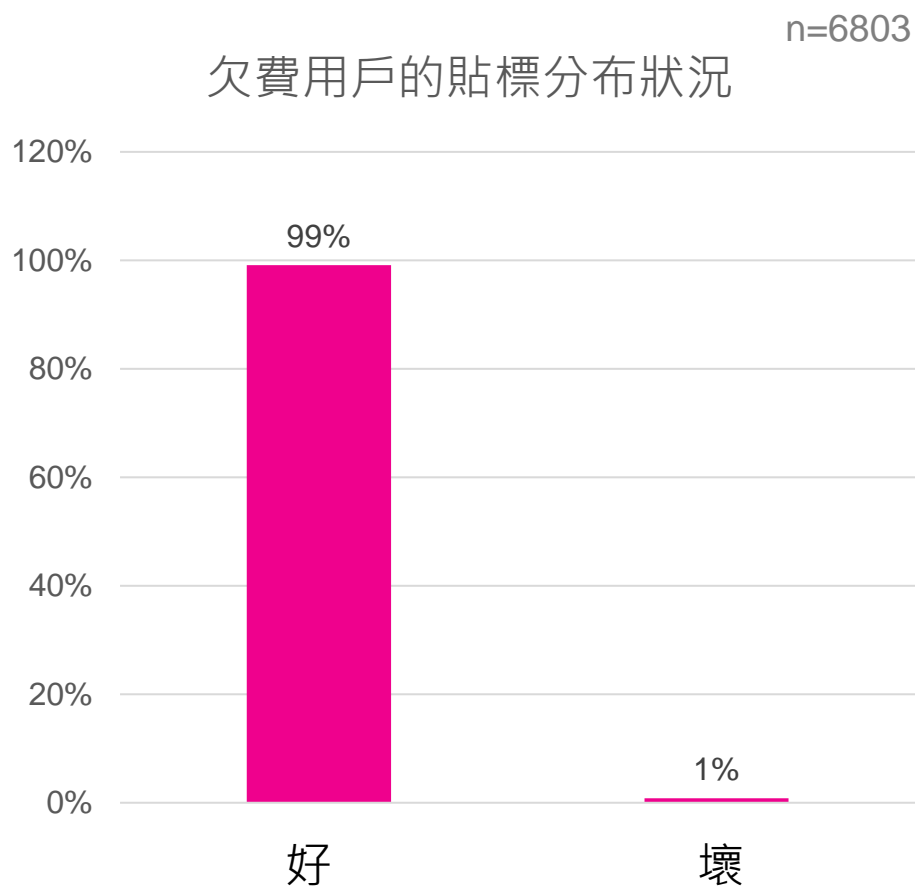
(1) RFM 研究 (2) LM & GLM 分析

結論

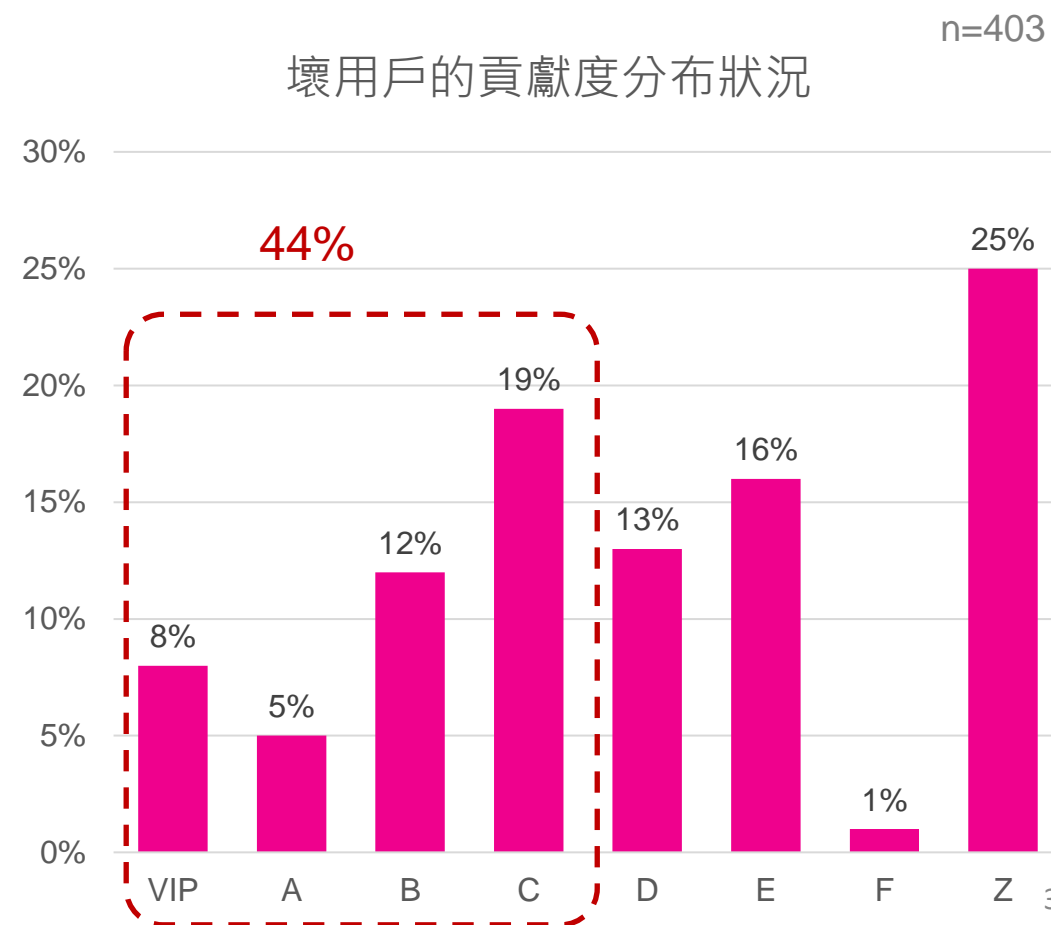
研究問題 |

目前的好/壞用戶分群(Flag 變數)並非針對DCB用戶，無法正確解釋用戶的繳費、貢獻度表現

欠費用戶只有1% 被貼標為壞用戶



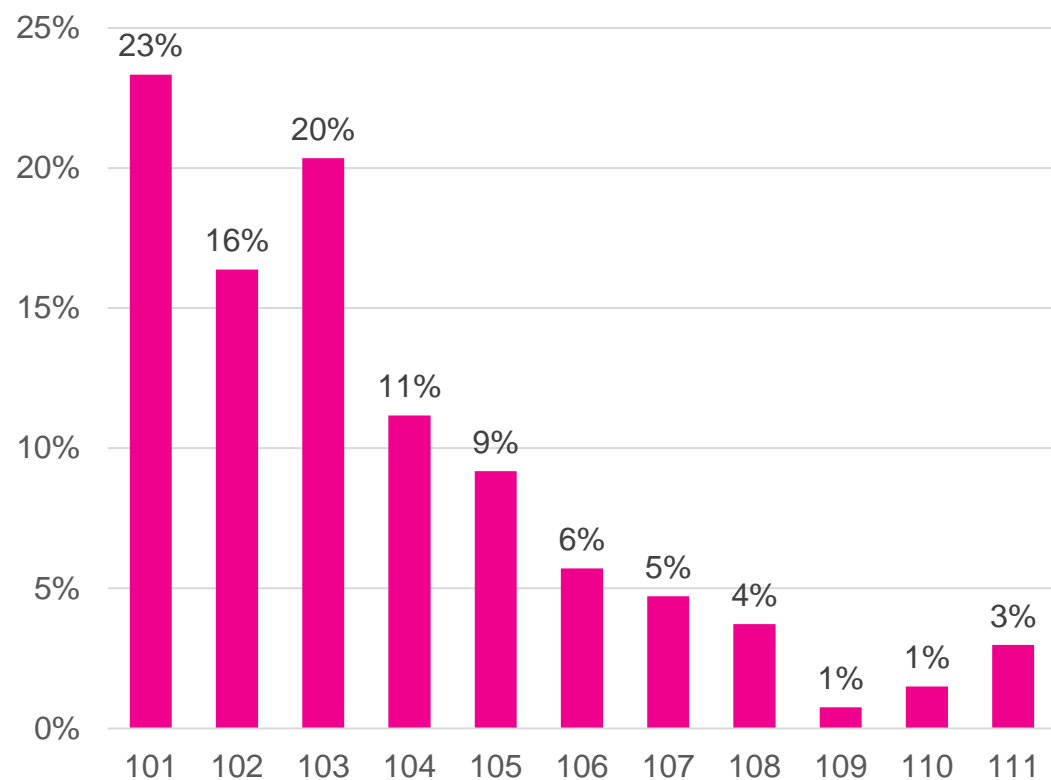
有44%的壞用戶為前段貢獻度用戶(VIP,A,B,C)



目前的好壞用戶分群的信用評等皆為較保守的分布

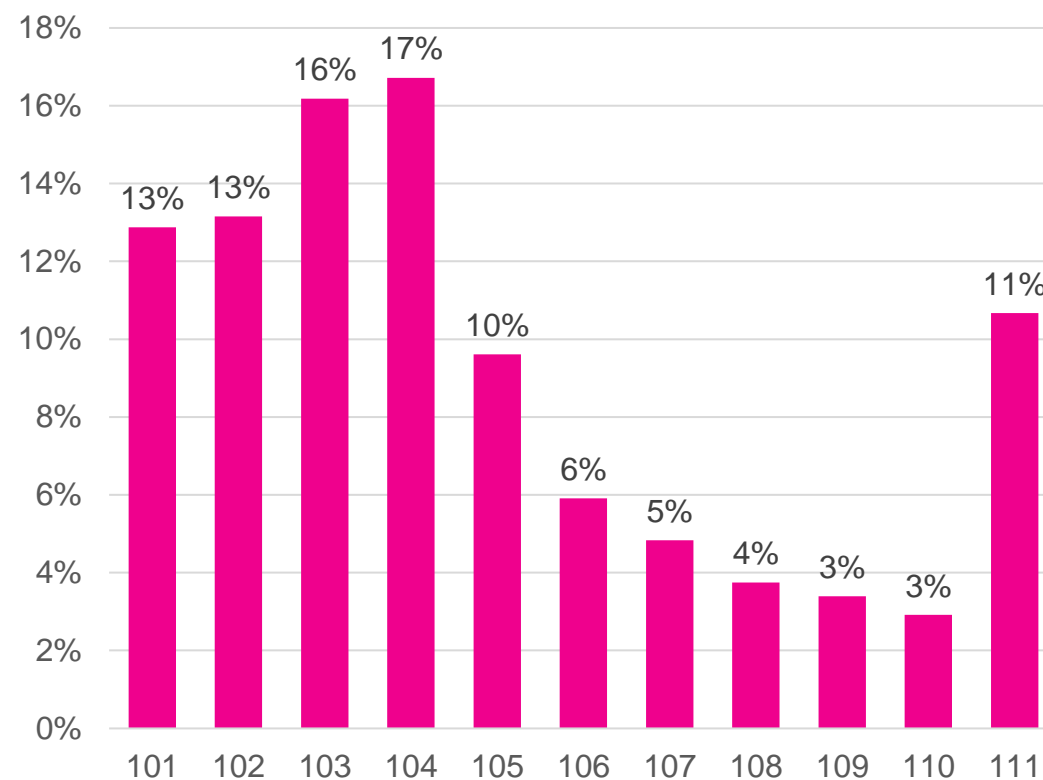
壞用戶的信用評等分布狀況

n=403



好用戶的信用評等分布狀況

n=32156



研究問題 |

目前分群較簡單，只有將用戶分成好/壞兩種。好用戶的信用評級偏向保守

25%

好用戶中近三個月曾經購買
失敗比例

12%

失敗金額約占信用等級比率

\$87

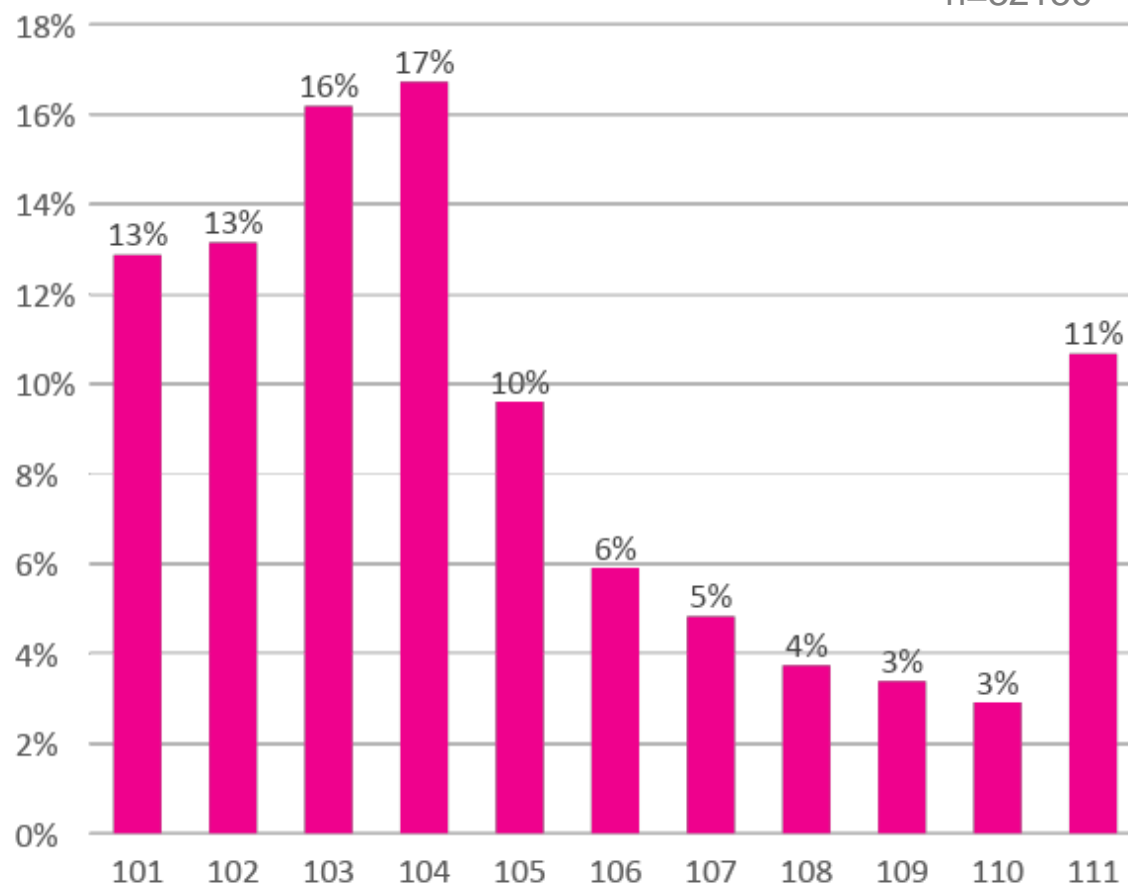
平均每月失敗金額

91%

此群用戶中，為近三個月整體
DCB用戶消費前50%的比例

好用戶的信用評等分布狀況

n=32156



建立以DCB為核心建立的用戶分群標準，以優化代收服務

RFM分析

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。

LM & GLM模型

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)

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。

RFM分析 |

定義 RFM 並將變數編碼，依照 M 指標定義用戶繳款能力

R

最近一次的消費天數

近 0 - 6 天	4
7 - 13 天	3
14 - 20 天	2
大於 20 天	1

F

三個月內消費三類品項的總次數

20 次以上	4
11 - 20 次	3
3 - 10 次	2
小於 3 次	1

M

近三個月平均繳費金額 -
近三個月的平均出帳金額

大於 0：繳費金額大於出帳金額
等於 0：無欠繳也無多繳
小於 0：有欠繳情形

3 高潛力用戶
2 低潛力用戶
1 壞用戶

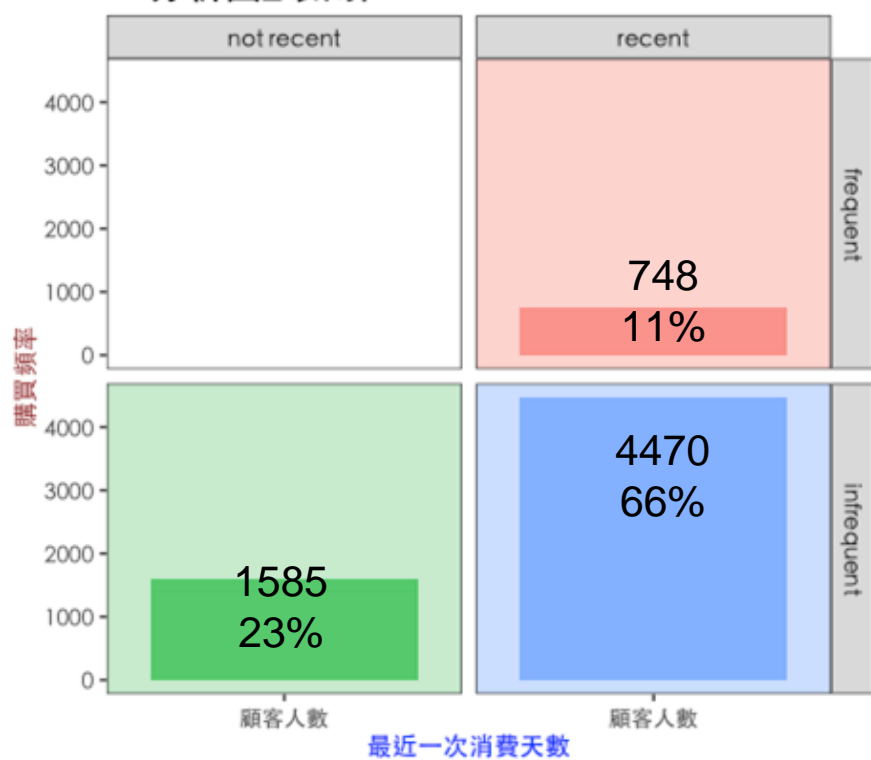
RFM分析 | 人數

高潛力用戶有52%為現活躍用戶，消費次數多的用戶會維持購買習慣

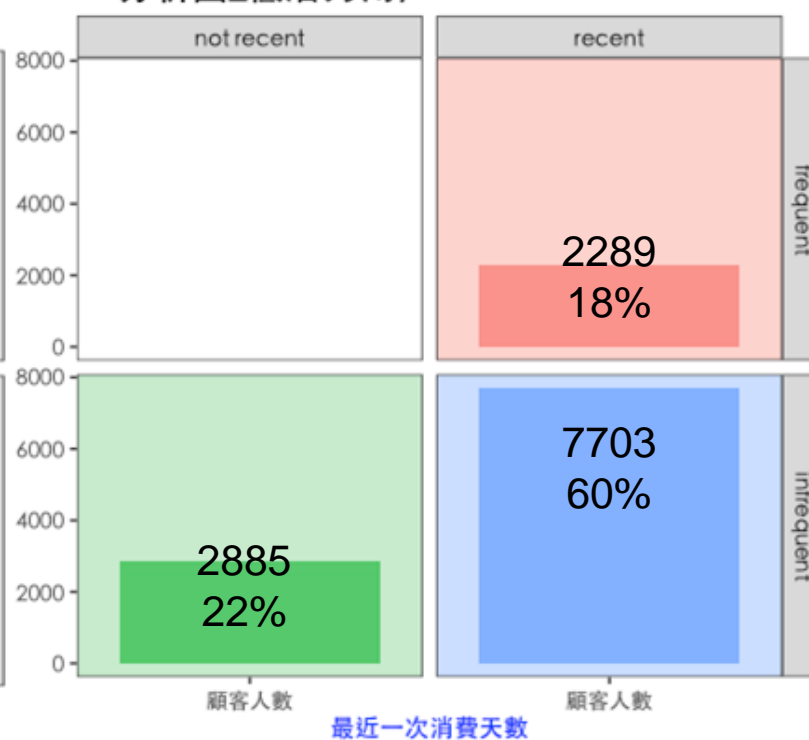
客群顏色指示表



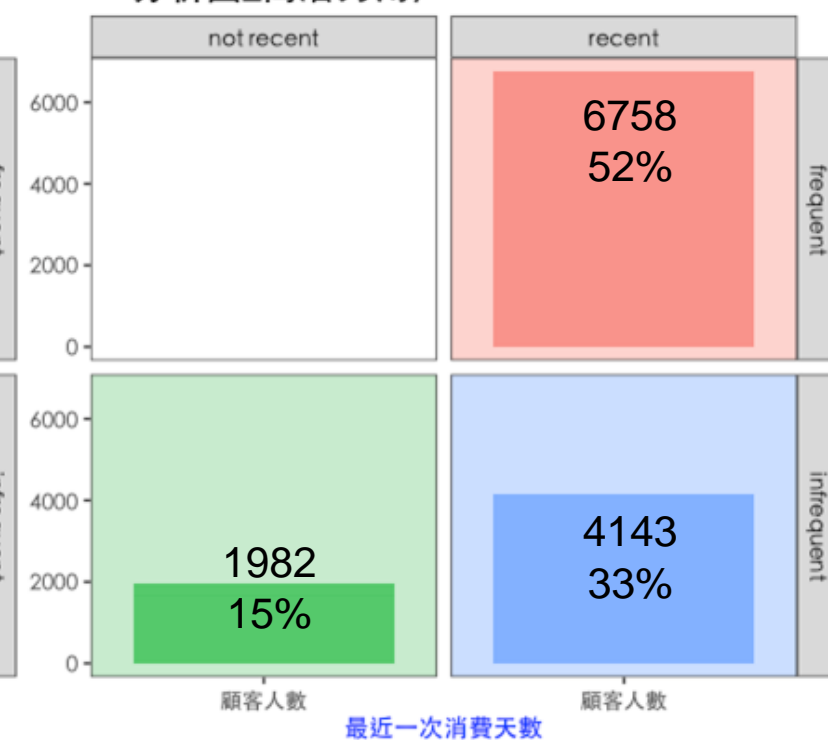
RF分析圖_壞用戶



RF分析圖_低潛力用戶



RF分析圖_高潛力用戶



RFM分析 | 性別

不論是哪個用戶群，現活躍用戶皆以男性比例較高

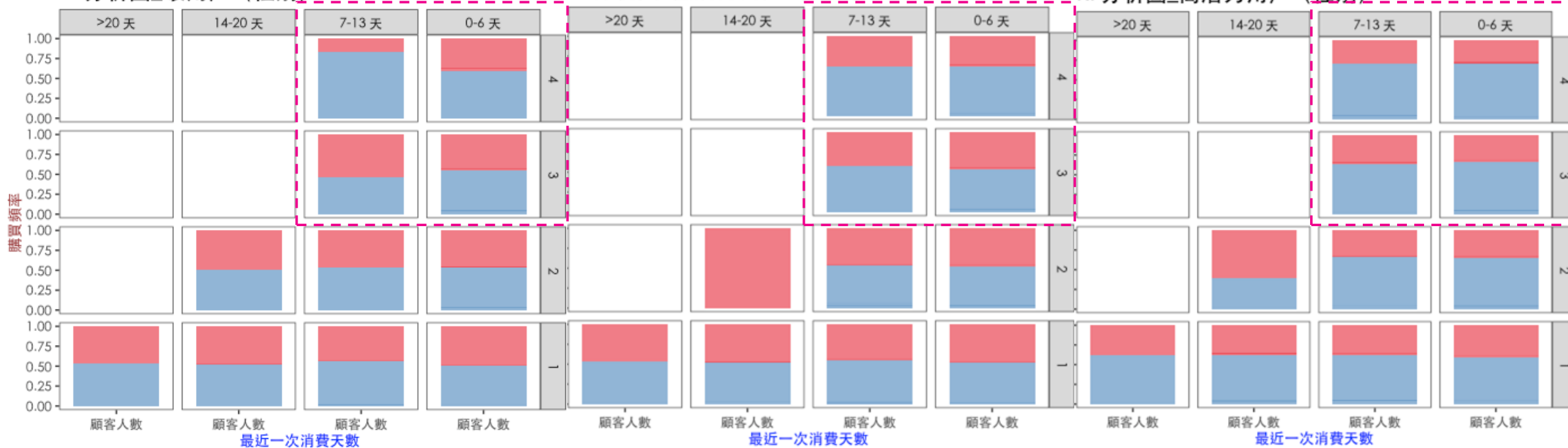
顧客性別



RF分析圖_壞用戶 (性別)

RF分析圖_低潛力用戶 (性別)

RF分析圖_高潛力用戶 (性別)



RFM分析 | 商品分類

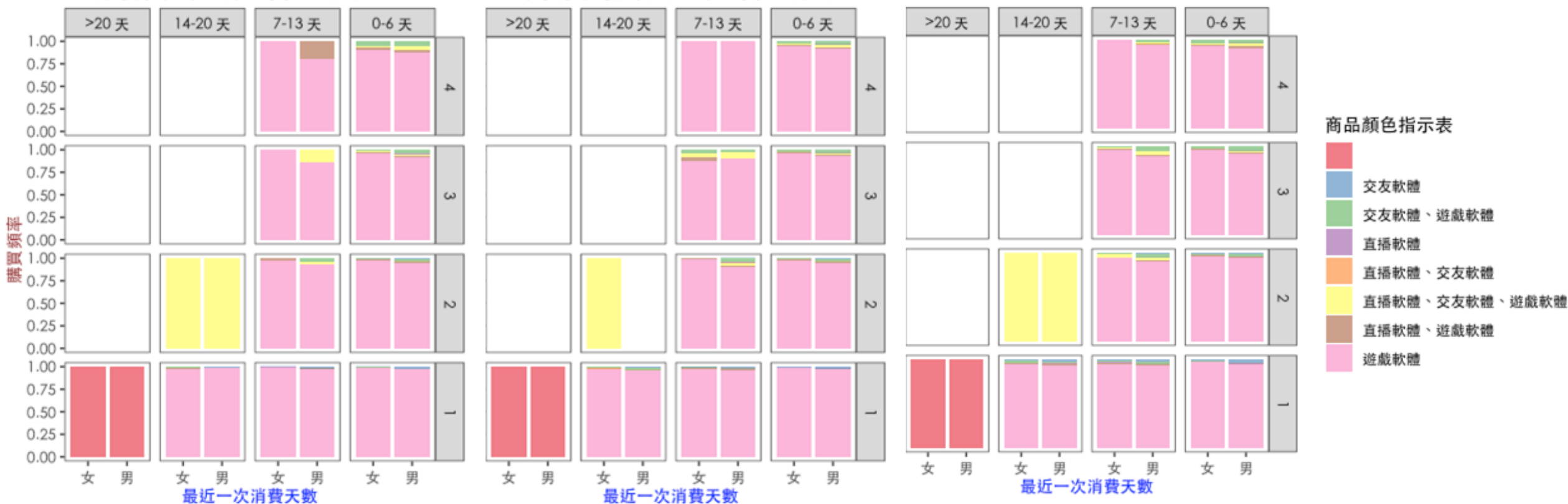
多數用戶以購買遊戲軟體為主；只買交友軟體的用戶購買頻率少

現活躍用戶中男性購買遊戲以外商品的比例較女性高

RF分析圖_壞用戶(商品分類)

RF分析圖_低潛力用戶(商品分類)

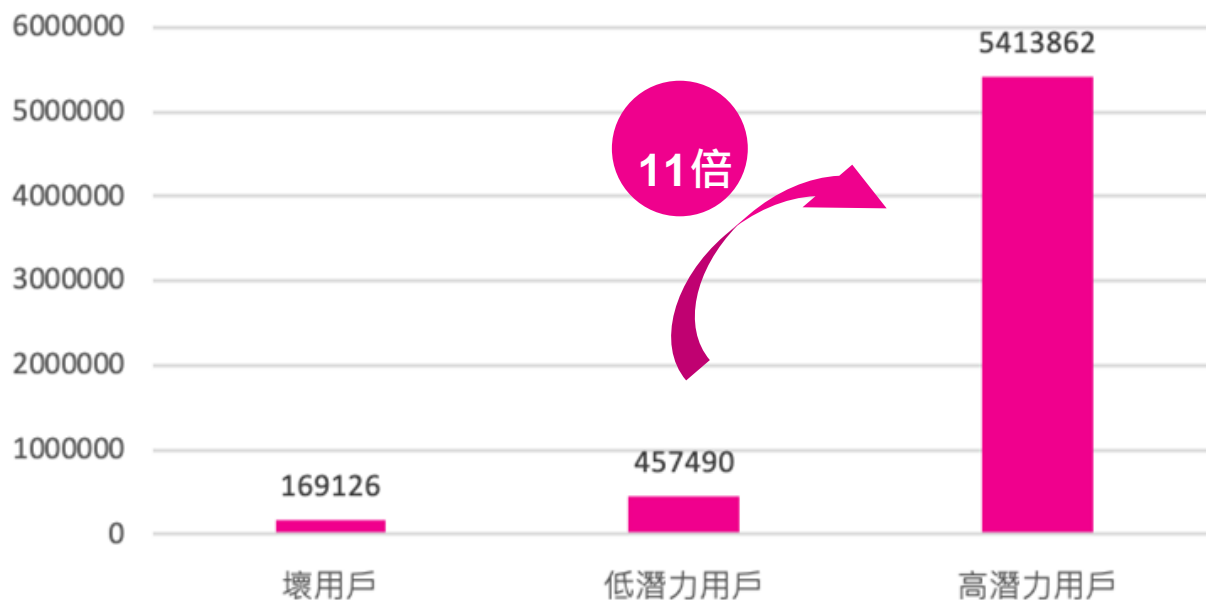
RF分析圖_高潛力用戶(商品分類)



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

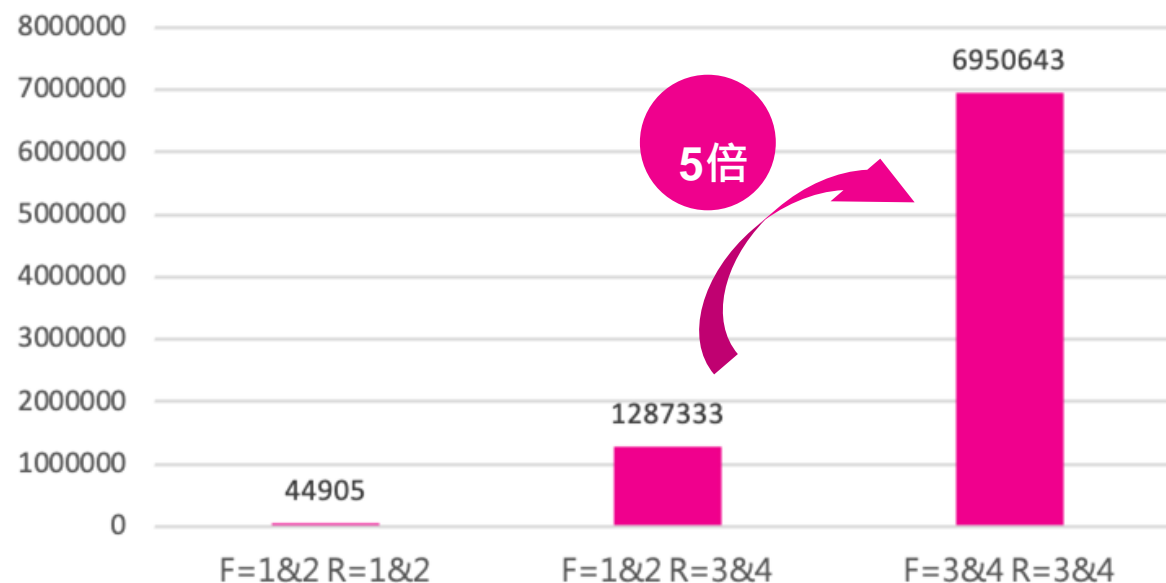
近期有消費行為且頻率高的
高潛力用戶的
消費金額龐大

近三個月購買總金額



現活躍的高潛力用戶之消費金額遠高於
消費頻率不高的嚐鮮用戶

高潛力用戶的總購買金額分佈



建議針對高潛力用戶群，提升 R 及 F 使嚐鮮用戶成為現活躍用戶

F = 1 & 2
R = 3 & 4

提升 F

- 提供 DCB 服務優惠，例如購買可提供電話費優惠

F = 1 & 2
R = 1 & 2

提升 R 及 F

- 針對上次購買的商品類型進行簡訊推播，提醒消費

RF分析圖_高潛力用戶(商品分類)



首先建構LM模型，以 **PAY** (近三個月每個月平均繳費金額 - 近三期平均出帳金額) 為反應變數Y，並搭配不同解釋變數建構兩模型：

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)

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。

首先建構LM模型，以 **PAY** (近三個月每個月平均繳費金額 - 近三期平均出帳金額) 為反應變數Y，並搭配不同解釋變數建構兩模型：

- 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)

以兩種方式初步建構線性模型：模型一

反應變數：PAY

解釋變數：

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)

CNTRB_LVL(出帳時用戶貢獻度)

HANDSET(手機廠牌)

N3_PAID_CHANNEL_CATE

(近三個月主要繳費通路)

N3_VAS_FEE_PER_MON

(近三個月平均增值服務消費金額)

N3_MSG_FEE_PER_MON

(近三個月平均每個月的簡訊費)

RATE_PLAN_RNG

(月底專案實收總資費級距)

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)

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)

L3_AVG_AMT_NITILE_OVER_50_FLAG

(是否為近三個月整體DCB用戶消費前50%)

以兩種方式初步建構線性模型：模型一

反應變數：PAY

解釋變數：

CUR_AGE(目前年齡)

TNR(累積租期)

用戶屬性

MAIN_MSISDN(主門號註記)

NP_OUT_GRP(新申裝或攜碼業者)

CNTRB_LVL(出帳時用戶貢獻度)

HANDSET(手機廠牌)

N3_PAID_CHANNEL_CATE

(近三個月主要繳費通路)

N3_VAS_FEE_PER_MON

(近三個月平均增值服務消費金額)

N3_MSG_FEE_PER_MON

(近三個月平均每個月的簡訊費)

RATE_PLAN_RNG

(月底專案實收總資費級距)

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)

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)

L3_AVG_AMT_NITILE_OVER_50_FLAG

(是否為近三個月整體DCB用戶消費前50%)

以兩種方式初步建構線性模型：模型一

反應變數：PAY

解釋變數：

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)

CNTRB_LVL(出帳時用戶貢獻度)

HANDSET(手機廠牌)

N3_PAID_CHANNEL_CATE

(近三個月主要繳費通路)

N3_VAS_FEE_PER_MON

(近三個月平均增值服務消費金額)

N3_MSG_FEE_PER_MON

(近三個月平均每個月的簡訊費)

繳款資訊

RATE_PLAN_RNG

(月底專案實收總資費級距)

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)

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)

L3_AVG_AMT_NITILE_OVER_50_FLAG

(是否為近三個月整體DCB用戶消費前50%)

以兩種方式初步建構線性模型：模型一

反應變數：PAY

解釋變數：

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)

CNTRB_LVL(出帳時用戶貢獻度)

HANDSET(手機廠牌)

N3_PAID_CHANNEL_CATE

(近三個月主要繳費通路)

N3_VAS_FEE_PER_MON

(近三個月平均增值服務消費金額)

N3_MSG_FEE_PER_MON

(近三個月平均每個月的簡訊費)

帳單細項資訊

RATE_PLAN_RNG

(月底專案實收總資費級距)

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)

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)

L3_AVG_AMT_NITILE_OVER_50_FLAG

(是否為近三個月整體DCB用戶消費前50%)

以兩種方式初步建構線性模型：模型一

反應變數：PAY

解釋變數：

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)

CNTRB_LVL(出帳時用戶貢獻度)

HANDSET(手機廠牌)

帳務資訊

N3_PAID_CHANNEL_CATE

(近三個月主要繳費通路)

N3_VAS_FEE_PER_MON

(近三個月平均增值服務消費金額)

N3_MSG_FEE_PER_MON

(近三個月平均每個月的簡訊費)

RATE_PLAN_RNG

(月底專案實收總資費級距)

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)

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)

L3_AVG_AMT_NITILE_OVER_50_FLAG

(是否為近三個月整體DCB用戶消費前50%)

以兩種方式初步建構線性模型：模型一

反應變數：PAY

解釋變數：

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)

CNTRB_LVL(出帳時用戶貢獻度)

HANDSET(手機廠牌)

N3_PAID_CHANNEL_CATE

(近三個月主要繳費通路)

N3_VAS_FEE_PER_MON

(近三個月平均增值服務消費金額)

N3_MSG_FEE_PER_MON

(近三個月平均每個月的簡訊費)

月底資訊

RATE_PLAN_RNG

(月底專案實收總資費級距)

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)

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)

L3_AVG_AMT_NITILE_OVER_50_FLAG

(是否為近三個月整體DCB用戶消費前50%)

以兩種方式初步建構線性模型：模型一

反應變數：PAY

解釋變數：

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)

CNTRB_LVL(出帳時用戶貢獻度)

HANDSET(手機廠牌)

N3_PAID_CHANNEL_CATE

(近三個月主要繳費通路)

N3_VAS_FEE_PER_MON

(近三個月平均增值服務消費金額)

N3_MSG_FEE_PER_MON

(近三個月平均每個月的簡訊費)

RATE_PLAN_RNG

(月底專案實收總資費級距)

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)

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)

L3_AVG_AMT_NITILE_OVER_50_FLAG

(是否為近三個月整體DCB用戶消費前50%)

DCB額度資訊

以兩種方式初步建構線性模型：模型一

反應變數：PAY

解釋變數：

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)

CNTRB_LVL(出帳時用戶貢獻度)

HANDSET(手機廠牌)

N3_PAID_CHANNEL_CATE

(近三個月主要繳費通路)

N3_VAS_FEE_PER_MON

(近三個月平均增值服務消費金額)

N3_MSG_FEE_PER_MON

(近三個月平均每個月的簡訊費)

RATE_PLAN_RNG

(月底專案實收總資費級距)

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)

DCB交易資訊

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)

L3_AVG_AMT_NITILE_OVER_50_FLAG

(是否為近三個月整體DCB用戶消費前50%)

	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_MSISDN	-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_LVL8	4.536e+00	2.109e+00	2.151	0.031510 *
CNTRB_LVL6	2.312e+00	2.302e+00	1.004	0.315292
CNTRB_LVL5	-1.257e+01	2.534e+00	-4.959	7.12e-07 ***
CNTRB_LVL4	-2.624e+01	2.651e+00	-9.897	< 2e-16 ***
CNTRB_LVL3	-9.530e+00	3.764e+00	-2.532	0.011353 *
CNTRB_LVL2	3.076e+00	2.259e+00	1.362	0.173352
CNTRB_LVL1	-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_FLAG2	-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_FLAG2	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

以兩種方式初步建構線性模型：模型一

模型一總結與詮釋：

1. 年齡層而言，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
2. 累積租期越長，對指標PAY亦有正面影響
3. 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。
4. 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。
5. 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。

以兩種方式初步建構線性模型：模型一

模型一總結與詮釋：

6. 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
7. 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
8. 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
9. 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
10. 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

以兩種方式初步建構線性模型：模型一

模型一總結與詮釋：

11. **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。
12. 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。

以兩種方式初步建構線性模型：模型二

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，建構出模型二：

反應變數：PAY

解釋變數：

大部分的模型一變數

G_N3_BUY_BUTTON50_CNT_CNT

G_N3_BUY_TOP5_CNT_CNT

近三個月購買前5% / 後50%熱門GOOGLE商品次數

G_N3_BUY_BUTTON50_CNT_OR_NOT

G_N3_BUY_TOP5_CNT_OR_NOT

近三個月是否購買前5% / 後50%熱門GOOGLE商品

G_N3_BUY_BUTTON50_CNT_AVG_AMT

G_N3_BUY_TOP5_CNT_AVG_AMT

近三個月是否購買前5% / 後50%GOOGLE商品平均金額

G_N3_BUY_BUTTON50_DCON_OR_NOT

G_N3_BUY_TOP5_DCON_OR_NOT

近三個月是否購買前5% / 後50%購買人次GOOGLE商品

G_N3_BUY_BUTTON50_AVG_PRICE_OR_NOT

G_N3_BUY_TOP5_AVG_PRICE_OR_NOT

近三個月購買後50% (前5%) 平均金額人次GOOGLE商品

G_N3_BUY_DATE_APP_OR_NOT

用戶近三個月是否購買交友軟體

G_N3_BUY_GAME_APP_OR_NOT

用戶近三個月是否購買遊戲軟體

N3_NORM_TIME_CNT_PER_MONTH

近三個月一般時刻平均每月成功消費金額

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_MSISDN	-1.184e+02	3.455e+00	-34.285	< 2e-16 ***
N3_OVER_PAY_FLAG	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_LVL	1.927e-01	2.612e+00	0.074	0.941193
CNTRB_LVL	5.043e-01	2.516e+00	0.200	0.841160
CNTRB_LVL	8.078e-01	2.620e+00	0.308	0.757834
CNTRB_LVL	-4.290e+00	2.547e+00	-1.685	0.092082 .
CNTRB_LVL	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_BOTTOM50_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_BOTTOM50_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_BOTTOM50_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_BOTTOM50_DCON_OR_NOTY	4.101e+00	3.049e+00	1.345	0.178623
G_N3_BUY_BOTTOM50_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

以兩種方式初步建構線性模型：模型二

模型二總結與詮釋：

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。

- a) 年齡區間：**20-29歲**的用戶PAY表現較其他區間好
- b) 付費通路：**直營門市付費**的用戶PAY表現較其他通路突出，

以兩種方式初步建構線性模型：模型二

模型二總結與詮釋：

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。

- a) 年齡區間：**20-29歲**的用戶PAY表現較其他區間好
- b) 付費通路：**直營門市付費**的用戶PAY表現較其他通路突出，

可以積極經營以上族群使用DCB服務

以兩種方式初步建構線性模型：模型二

模型二總結與詮釋：

比較特別的發現

- a) 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
- b) 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

以兩種方式初步建構線性模型：模型二

模型二總結與詮釋：

比較特別的發現

- a) 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
- b) 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

以上顯示培養使用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

修改新產生的衍生變數 M

將原先 M 編碼為 1 修改為 0

將原先 M 編碼為 2 與 3 修改為 1

涵蓋的即是「壞用戶」(原先指標 $PAY < 0$ 者)

涵蓋的即是「高潛力用戶」與「低潛力用戶」
(原先指標 $PAY \geq 0$ 者)

修改新產生的衍生變數 M

將原先 M 編碼為 1 修改為 0

將原先 M 編碼為 2 與 3 修改為 1

涵蓋的即是「壞用戶」(原先指標 $PAY < 0$ 者)

涵蓋的即是「高潛力用戶」與「低潛力用戶」
(原先指標 $PAY \geq 0$ 者)

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

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)

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	***

GLM分析 | 從LM到GLM

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_MSISDN	-2.387e+00	1.895e-01	-12.599	< 2e-16	***
G_N3_BUY_BUTTON50_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_BUTTON50_CNT_OR_NOT	-4.577e-01	4.389e-01	-1.043	0.297021	
G_N3_BUY_BUTTON50_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_BUTTON50_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_BUTTON50_DCON_OR_NOT	NA	NA	NA	NA	
G_N3_BUY_BUTTON50_DCON_OR_NOTY	1.732e-01	1.379e-01	1.256	0.209068	
G_N3_BUY_BUTTON50_AVG_PRICE_OR_NOT	NA	NA	NA	NA	
G_N3_BUY_BUTTON50_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

GLM分析 | 從LM到GLM

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_MSISDN	-2.387e+00	1.895e-01	-12.599	< 2e-16	***
G_N3_BUY_BUTTON50_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_BUTTON50_CNT_OR_NOT	-4.577e-01	4.389e-01	-1.043	0.297021	
G_N3_BUY_BUTTON50_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_BUTTON50_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_BUTTON50_DCON_OR_NOT	NA	NA	NA	NA	
G_N3_BUY_BUTTON50_DCON_OR_NOTY	1.732e-01	1.379e-01	1.256	0.209068	
G_N3_BUY_BUTTON50_AVG_PRICE_OR_NOT	NA	NA	NA	NA	
G_N3_BUY_BUTTON50_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

此模型中 deviance 從 33378 至 18032

Number of Fisher Scoring iterations: 8

DCB變數

- 加入DCB變數對Deviance的影響很小，其中一個可能是DCB變數的資料量太少所致，故決定不採納DCB相關變數

多元
共線性

- 利用GVIF計算作GLM下多元共線性之檢查
- 當VIF的閾值訂為10，而 $GVIF^{1/(2*df)}$ 閾值則會是 $\sqrt{10}=3.16$
- 捨去GVIF過大者再以剩下的變數計算後，最後應無多元共線性問題

	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

- 連續變數：如 **BILL_AMT**，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 ->

$$\exp(0.009342 \times (200 - 100)) = 2.55$$

- 類別變數：如 **MAIN_MSISDN**，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 ->

$$\exp(-2.028) = 0.13$$

```
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_LVL B	2.245e-01	1.170e-01	1.920	0.054912 .
CNTRB_LVL C	2.325e-01	1.119e-01	2.077	0.037807 *
CNTRB_LVL D	-8.469e-02	1.152e-01	-0.735	0.462357
CNTRB_LVL E	-6.436e-01	1.153e-01	-5.584	2.35e-08 ***
CNTRB_LVL F	2.089e-01	1.392e-01	1.501	0.133378
CNTRB_LVL VIP	-3.140e-01	1.460e-01	-2.151	0.031484 *
CNTRB_LVL Z	-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_MSISDN Y	-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

- 當模型上的解釋變數係數為正值，對於我們所定義「具潛力用戶」指標有正面影響；解釋變數係數為負值時則反之
- 若匯入一筆真實資料至模型內，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 (threshold) 為歸類二元變數的基準
- 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」



根據這個方法優化代收的分群基準

問題：現有的Flag分群非針對代收服務，無法精確反映用戶表現

目標：建立以DCB為核心建立的用戶分群標準，並藉由分析用戶行為提高DCB服務使用，來優化代收服務

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送，提醒消費

LM & 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