

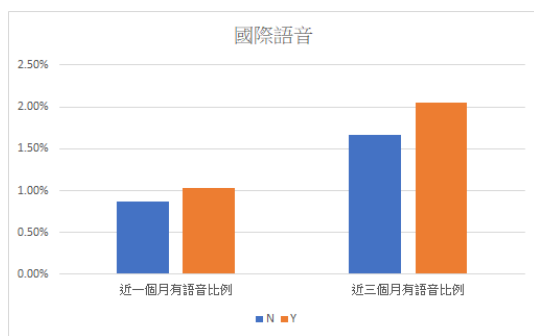
### 第三組：電信代收優化

- 研究目標：用戶分群再定義一好、壞、高潛力用戶，以優化代收服務
- 分群基準：溢繳、信用評級
- EDA研究方法：初步選定以下六個主題，與分群基準對照，探討影響分群可能的變數

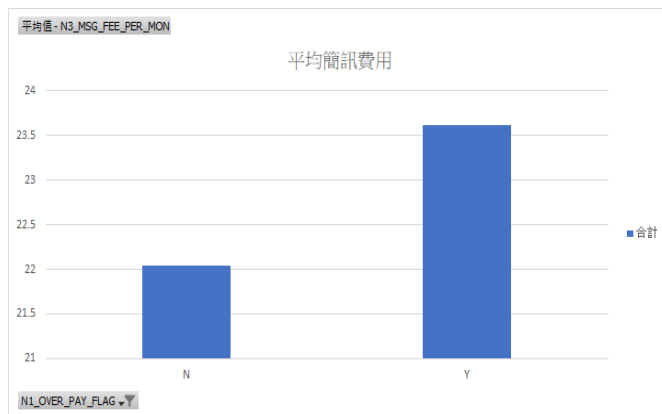
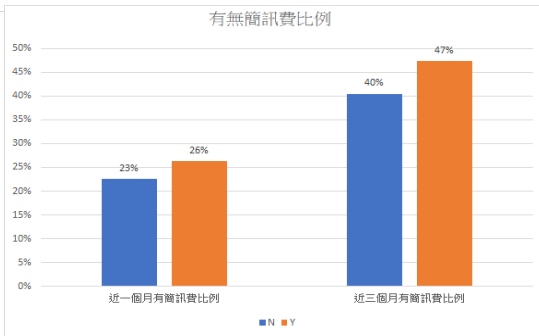
#### 一、有無超值服務、國際語音、簡訊費與分群基準的關係

##### 分群基準：溢繳

-有無國際語音



-有無簡訊費比例



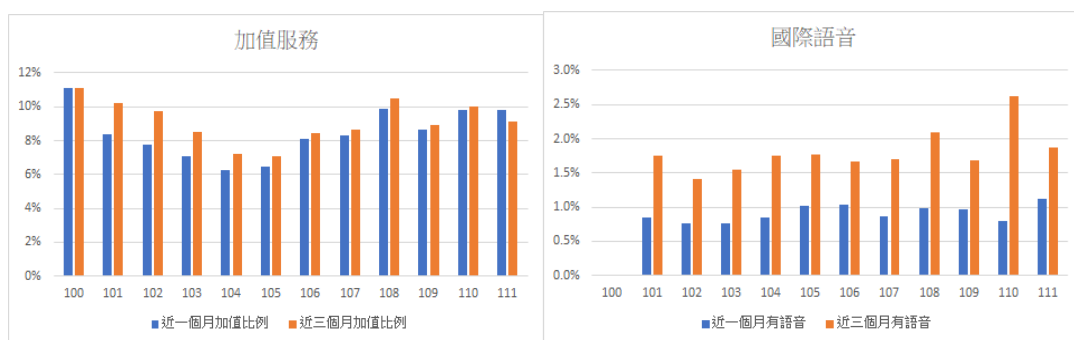
根據上圖，可得以下觀察：

- 有無溢繳的用戶使用超值服務的比例差不多
- 有溢繳的人有國際語音的比例略高
- 有溢繳的人有簡訊費的比例略高，平均簡訊費用約高出7%

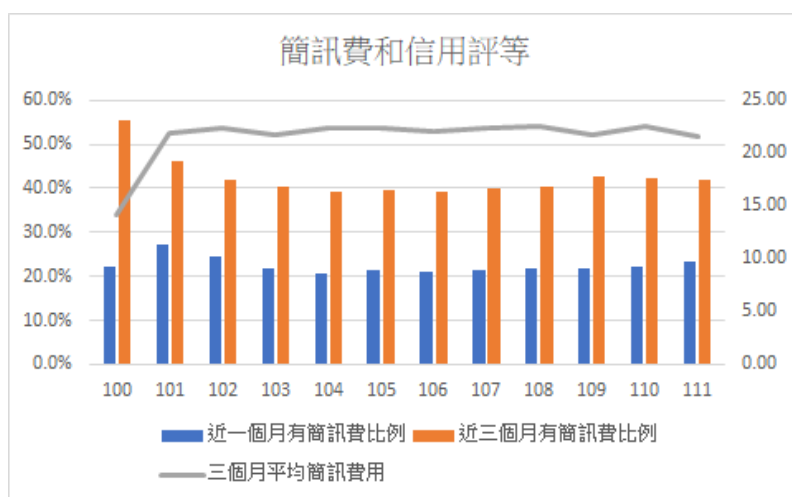
##### 分群基準：信用評級

-有無使用超值服務

-有無國際語音



### -有無簡訊費比例

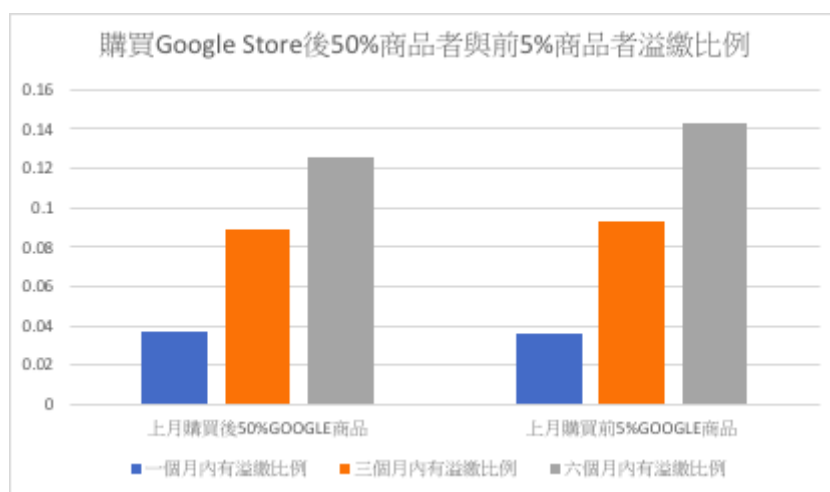


根據上圖，可得以下觀察：

- 有無使用加值服務在每個信用評級的比例有差異，略為呈現 2 次相關的趨勢(信用評級高及信用評級低的使用比例較多)。用戶一個月及三個月的行為接近，呈現正相關。
- 有無國際語音在每個信用評級的比例略有差異，但看不出相關性。
- 除了信用評級為 100 以外，其他信用評級在有無簡訊費的比例沒有顯著差異。信用評級為100的用戶近一個月有簡訊費的比例較低、近三個月的有簡訊費的比例較高。信用評級為 100 的平均簡訊費用最低，其他則無顯著差距。

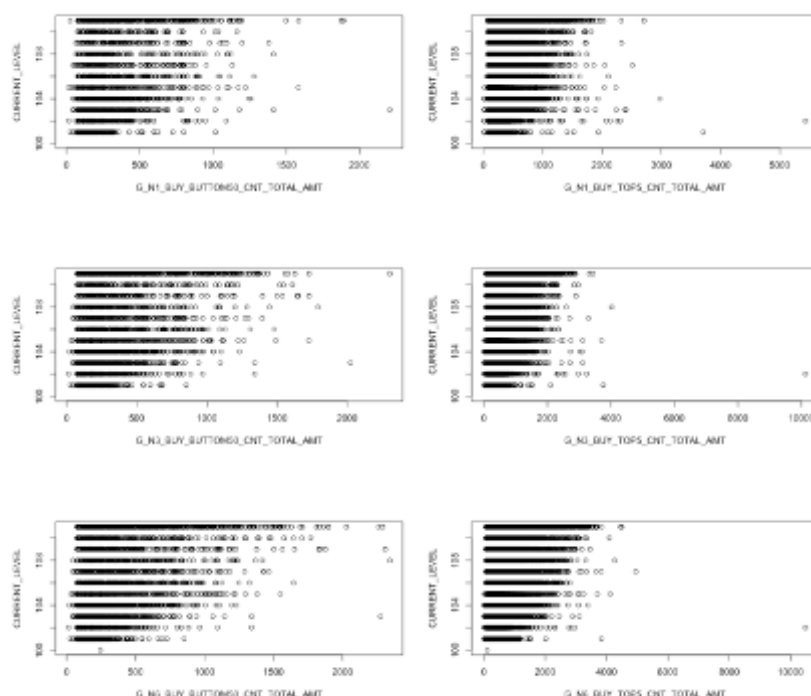
## 二、購買Google store後50%與前5%商品與分群基準的關係

分群基準：溢繳



上圖針對「上月有購買後50%商品」者與「上月有購買前5%商品者」分別觀察其一個月內、三個月內、六個月內有溢繳的比例。結果發現「上月有購買後50%商品」者與「上月有購買前5%商品」者一個月內有溢繳的比例無太大差異，而三個月內有溢繳的比例與六個月內有溢繳的比例皆以「上月有購買前5%商品者」為高。

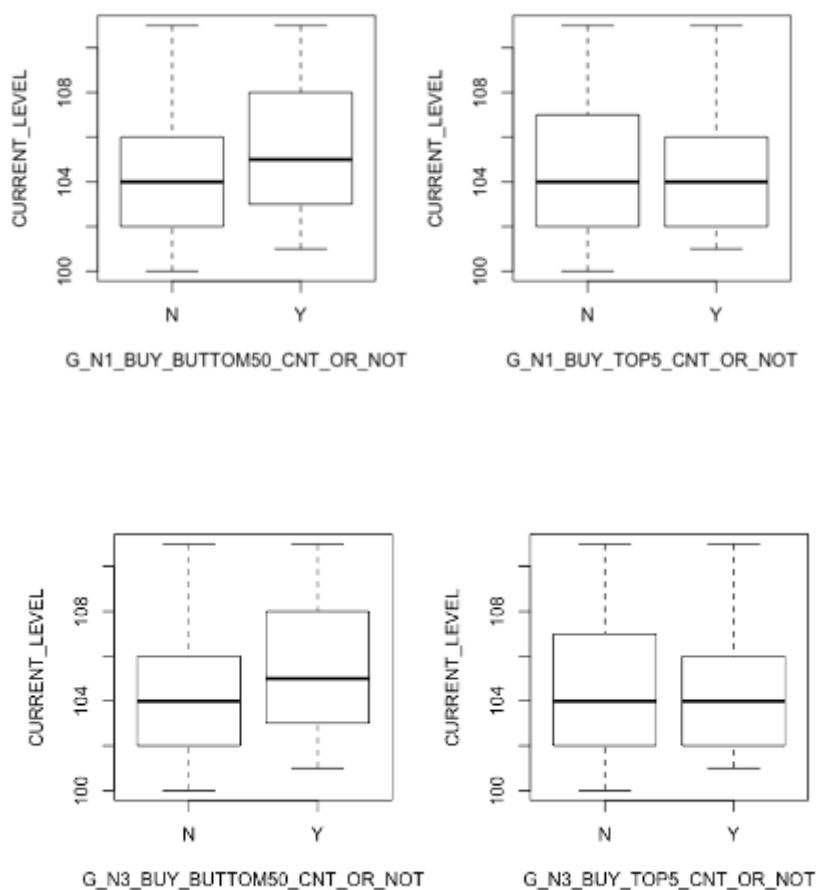
## 分群基準：信用評等



從一個月內、三個月內、六個月內Google後50%與前5%分別消費總額與信用評等的散佈圖來看，整體而言，無論是購買後50%或前5%者，在消費總額較低的情形下用戶信用評級高低皆有；而消費後50%總金額較高者普遍有信用評等較高，消費前5%總金額較高者亦有類似現象，

但相對而言較不明顯，另外，也發現有消費前5%商品金額極高而信用評等低的用戶出現在散布圖上。

有無購買後50%商品與前5%商品與信用評等之關係：



無論是以一個月或三個月內有無購買後50%商品或前5%商品的盒狀圖來看，有購買後50%商品的用戶之信用評等中位數皆較無購買者高；而有購買前5%商品的用戶之信用評等與無購買者之信用評等中位數則沒有明顯差別。

### 三、平均DCB購買次數、平均購買金額與分群基準的關係

首先觀察平均購買次數與金額，以區間方式來呈現其人數分佈，可以發現購買次數在10-20次內的比例高達46.2%，而金額則多落在100-300元內。

次數區間 人數 佔比	近一個月平均DCB購買次數區間人數							
	0	1-10	10-20	20-30	30-40	40-50	50-100	100以上
	2121	6656	23217	8420	5232	2061	2369	162
	4.2%	13.2%	46.2%	16.8%	10.4%	4.1%	4.7%	0.3%

近一個月平均DCB購買金額區間人數								
金額區間	0	1-100	100-200	200-300	300-400	400-500	500-1000	1000以上
人數	2122	7582	13511	9042	5482	3916	6907	1686
佔比	4.2%	15.1%	26.9%	18.0%	10.9%	7.8%	13.7%	3.4%

### 分群基準：溢繳

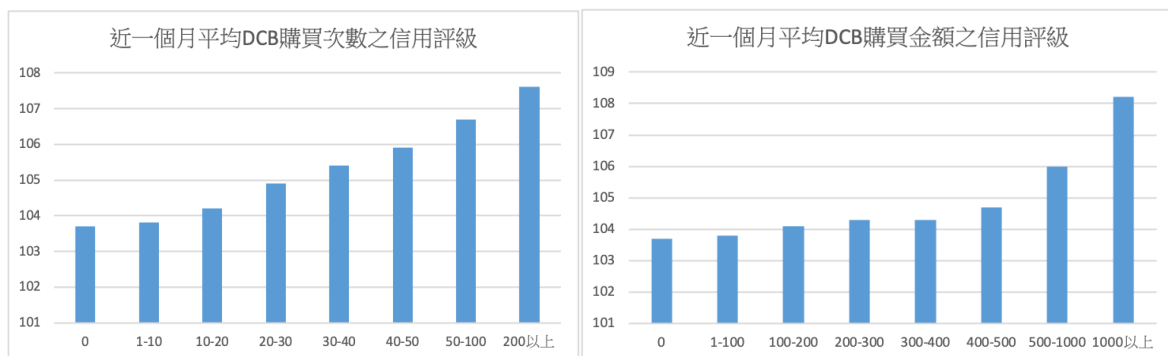
接著在各個購買次數與金額區間內，觀察溢繳的人數和比例，可知雖然在次數和金額高的區間內，溢繳比例都是最高的，不過整體而言差距並不大。

近一個月平均DCB購買次數溢繳比例								
次數區間	0	1-10	10-20	20-30	30-40	40-50	50-100	100以上
溢繳人數	74	219	774	355	194	96	125	10
溢繳比例	3.5%	3.3%	3.3%	4.2%	3.7%	4.7%	5.3%	6.2%

近一個月平均DCB購買金額溢繳比例								
金額區間	0	1-100	100-200	200-300	300-400	400-500	500-1000	1000以上
溢繳人數	73	230	437	352	242	147	270	92
溢繳比例	3.4%	3.0%	3.2%	3.9%	4.4%	3.8%	3.9%	5.5%

### 分群基準：信用評級

觀察不同次數與金額區間內，用戶的平均信用評級，可以發現兩者皆呈現遞增趨勢，也就是說隨著購買次數與金額增加，用戶的平均信用評級也越高。



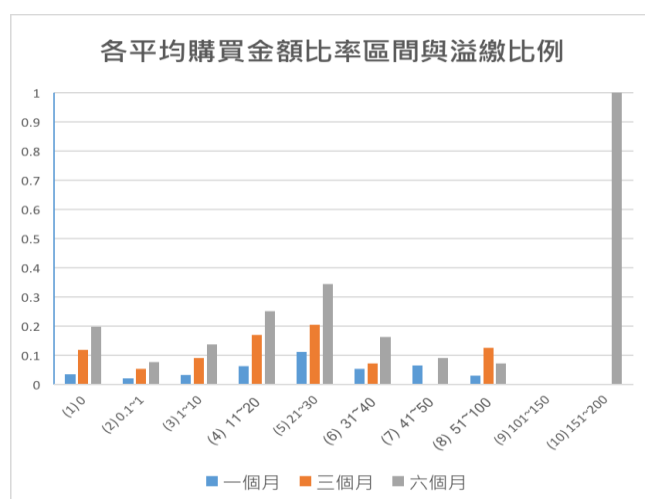
## 四、平均購買金額佔信用等級比率與分群基準的關係

以下分別為近一個月、三個月、六個月的平均購買金額佔信用等級比率區間人數表，可觀察到用戶主要集中在比例1~10，其次為11~20。

各平均購買金額比率區間人數										
	(1) 0	(2) 0.1~1	(3) 1~10	(4) 11~20	(5) 21~30	(6) 31~40	(7) 41~50	(8) 51~100	(9) 101~150	(10) 151~200
近一個月	2003	1793	37273	8497	332	93	31	33	7	2
近三個月	581	1632	41076	6492	210	42	13	16	1	1
近六個月	207	1261	41851	6323	233	37	11	14	1	1

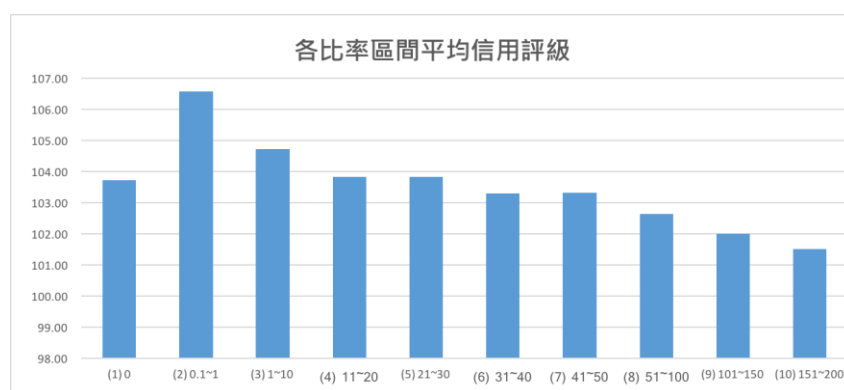
### 分群基準：溢繳

以下為近一個月、三個月、六個月的平均購買金額佔信用等級比率區間與溢繳情形的比例圖，除了51以上的區間數量少、情況較特殊外，大多數區間的溢繳情形逐漸下降，且由21-30區間表現最為突出。

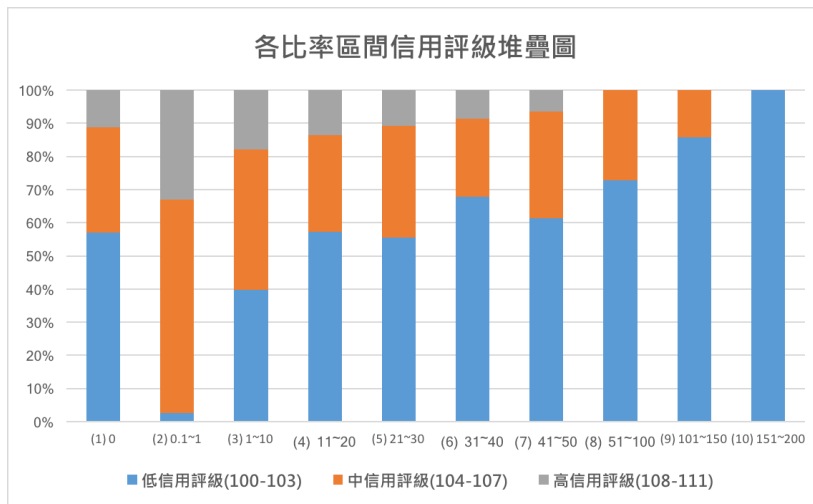


## 分群基準：信用評級

以近一個月各平均購買金額佔信用等級比率觀察各區間平均信用評級，可以發現0.1-1區間為目前評級最高，並有逐漸下降的趨勢，且大多數區間的平均評級落在低信用評級(100-103)層級。



以下將信用評級拆成低(100-103)、中(104-107)、高(108-111)三個層級，與近一個月各平均購買金額佔信用等級比率區間做比較：



根據(四)所有圖表與觀察，可得以下論點：

- 在0.1-1的區間，信用評級的表現較其他區間良好許多，且隨區間排序，低評級組成比例越大，推斷目前的評級方式是為保守的。而大多數人位於1-10區間，因此可藉由重新分群，增加獲益的空間。
- 對照所有區間於溢繳情形以及信用評等間的關係，能發現溢繳情況突出(11-20區間&21-30區間)的區間與其目前的信用評級平均並不相符，認為可進一步探究此兩個區間在使用電信服務上的表現，進一步分析發揮潛力的機會點。

## 五、月底資訊變數與分群基準的關係

### 月底資訊變數：

月底專案實收總資費級距(7階)	RATE_PLAN_RNG
月底專案原訂語音數據資費	MONTHLY_RATE_S
月底專案原訂數據資費	MONTHLY_RATE_I
月底專案原訂語音資費	MONTHLY_RATE_V
月底專案原定總資費	TOTAL_MONTHLY_RATE
月底專案實收語音數據資費	MONTHLY_FEE_S
月底專案實收數據資費	MONTHLY_FEE_I
月底專案實收語音資費	MONTHLY_FEE_V
月底專案實收總資費	TOTAL_MONTHLY_FEE
月底專案語音吃到飽	VOICE_UNLIM

### 分群變數：

是否為壞用戶	FLAG
近一個月是否曾欠費	N1_OWE_MNY_FLAG

近一個月是否曾溢繳

N1\_OVER\_PAY\_FLAG

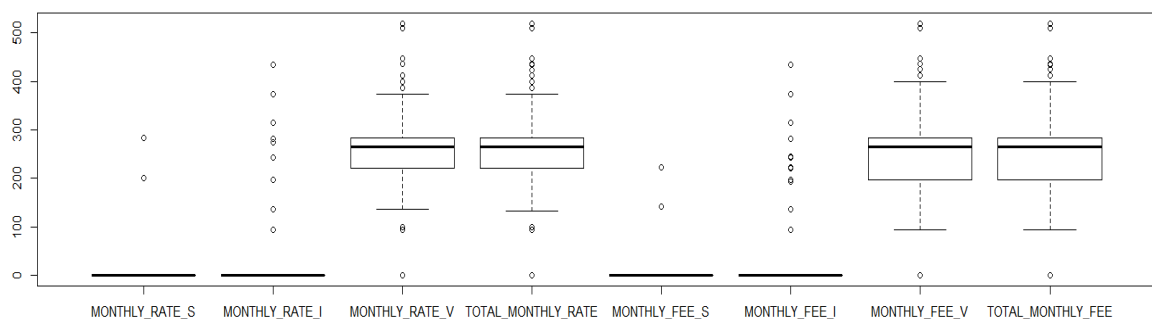
信用評級

CURRENT\_LEVEL

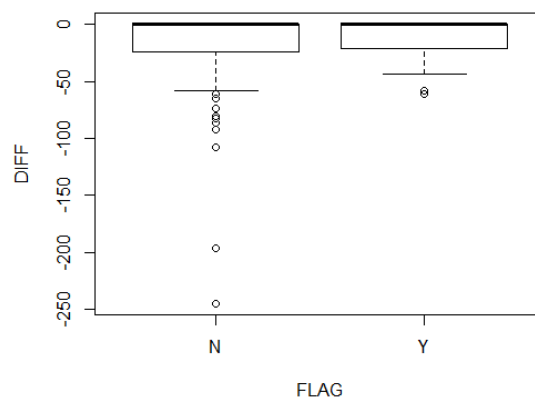
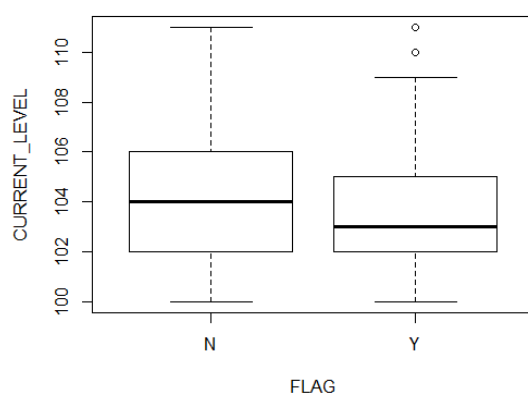
衍生變數：

月底專案實收總資費 - 月底專案原定總資費      DIFF

1. 使用「原訂」和「實收」的差距作為變數的原因是我認為最後實際收到的錢和預計要收的錢的差距會影響用戶分群。
2. 針對不同資費數據和「原訂」和「實收」兩類比較，繪製boxplot。
  - a. MONTHLY\_RATE\_S, MONTHLY\_RATE\_I, MONTHLY\_FEE\_S和MONTHLY\_FEE\_I 大部分數據都為0，只有少部分數據大於0。
  - b. MONTHLY\_RATE\_V, TOTAL\_MONTHLY\_RATE, MONTHLY\_FEE\_V和TOTAL\_MONTHLY\_FEE大部分數據集中在150和350之間。



3. 分別針對FLAG和CURRENT\_LEVEL, DIFF做Boxplot
  - a. 壞用戶(FLAG=Y)的信用評級較低。
  - b. 「原訂」和「實收」金額差異對於好壞用戶差異不大，因為差距大部分是0。



4. Summary report



- MONTHLY\_RATE\_S, MONTHLY\_RATE\_I, MONTHLY\_FEE\_S和MONTHLY\_FEE\_I有超過3/4的數據為0。
- MONTHLY\_RATE\_V, TOTAL\_MONTHLY\_RATE, MONTHLY\_FEE\_V和TOTAL\_MONTHLY\_FEE大部分數據集中在150和350之間。
- VOICE\_UNLIM 數據超過98%都是網內語音吃到飽。
- N1\_OWE\_MNY\_FLAG 數據全是沒有(N)。
- N1\_OVER\_PAY\_FLAG 數據超過96%都是沒有(N)。
- CURRENT\_LEVEL分布在100到111之間。
- DIFF有超過一半的數據都是0，也就是「原訂」和「實收」金額一樣。
- FLAG為Y的比例約為1%，也就是壞用戶比例約為1%。

```

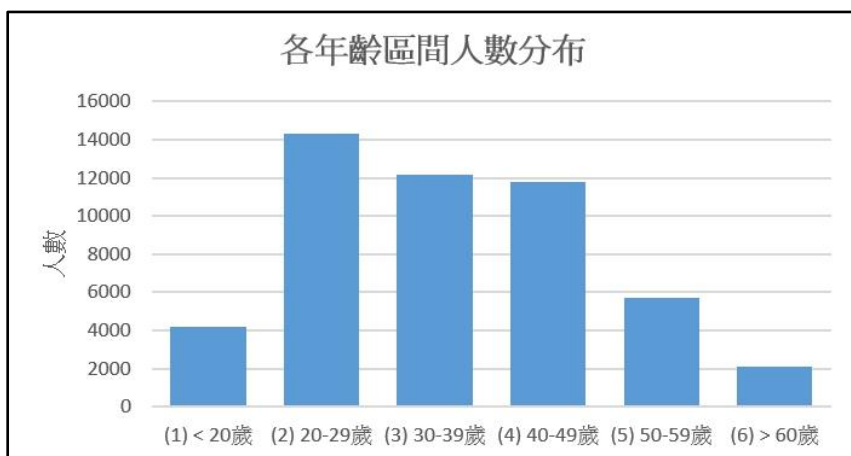
MONTHLY_RATE_S  MONTHLY_RATE_I  MONTHLY_RATE_V  TOTAL_MONTHLY_RATE  MONTHLY_FEE_S  MONTHLY_FEE_I  MONTHLY_FEE_V
Min. : 0.0000    Min. : 0.0000    Min. : 0.0      Min. : 0.0          Min. : 0.0000    Min. : 0.0000    Min. : 0.0
1st Qu.: 0.0000    1st Qu.: 0.0000    1st Qu.:220.9    1st Qu.:220.9        1st Qu.: 0.0000    1st Qu.: 0.0000    1st Qu.:197.0
Median : 0.0000    Median : 0.0000    Median :264.4    Median :264.4        Median : 0.0000    Median : 0.0000    Median :264.4
Mean : 0.1784     Mean : 0.4738     Mean :258.5     Mean :261.5         Mean : 0.1406     Mean : 0.4605     Mean :250.0
3rd Qu.: 0.0000    3rd Qu.: 0.0000    3rd Qu.:282.7    3rd Qu.:282.7        3rd Qu.: 0.0000    3rd Qu.: 0.0000    3rd Qu.:282.7
Max. :282.7000     Max. :434.5000    Max. :519.5     Max. :519.5         Max. :223.4000    Max. :434.5000    Max. :519.5
TOTAL_MONTHLY_FEE  VOICE_UNLIM  RATE_PLAN_RNG  FLAG  N1_OWE_MNY_FLAG  N1_OVER_PAY_FLAG  CURRENT_LEVEL
Min. : 0.0      未定義      : 15      (1) LL : 0~200 : 3214  N:49722  N:50240  N:48400  Min. :100.0
1st Qu.:197.0    網內語音吃到飽:49503  (2) L : 201~500 :13709  Y: 518      Y: 1840      1st Qu.:102.0
Median :264.4    網內語音計費 : 722    (3) ML : 501~700 :10511  (4) M : 701~900 :14564  Median :104.0
Mean :254.5      (5) MH : 901 以上: 8242  Mean :104.6
3rd Qu.:282.7      3rd Qu.:106.0
Max. :519.5      Max. :111.0
DIFF
Min. : -244.700
1st Qu.: -23.900
Median : 0.000
Mean : -6.954
3rd Qu.: 0.000
Max. : 0.000

```

## 5. 結論與推論

- 因為N1\_OWE\_MNY\_FLAG數據全是沒有(N)，所以未來分析要去掉這個變數。
- 因為MONTHLY\_RATE\_S, MONTHLY\_RATE\_I, MONTHLY\_FEE\_S和MONTHLY\_FEE\_I 大部分數據都為0，所以未來分析可能可以不用列入考慮。
- DIFF分布對於FLAG的影響有限。

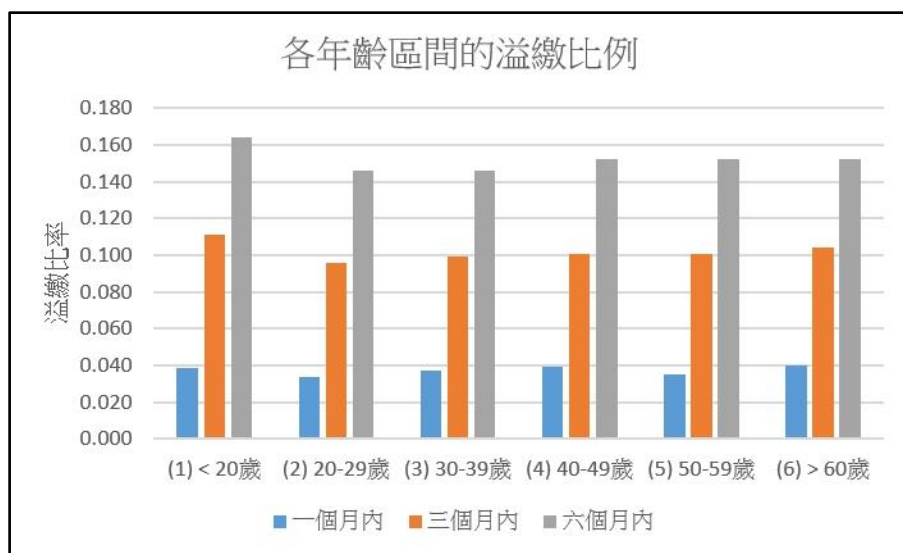
## 六、年齡與分群基準的關係



以下是所有用戶的年齡分布圖，可觀察到主要年齡為20-49歲，其中又以20-29歲為重。

### 分群基準：溢繳

以下為各年齡區間溢繳比率圖



根據上圖，可得以下觀察：

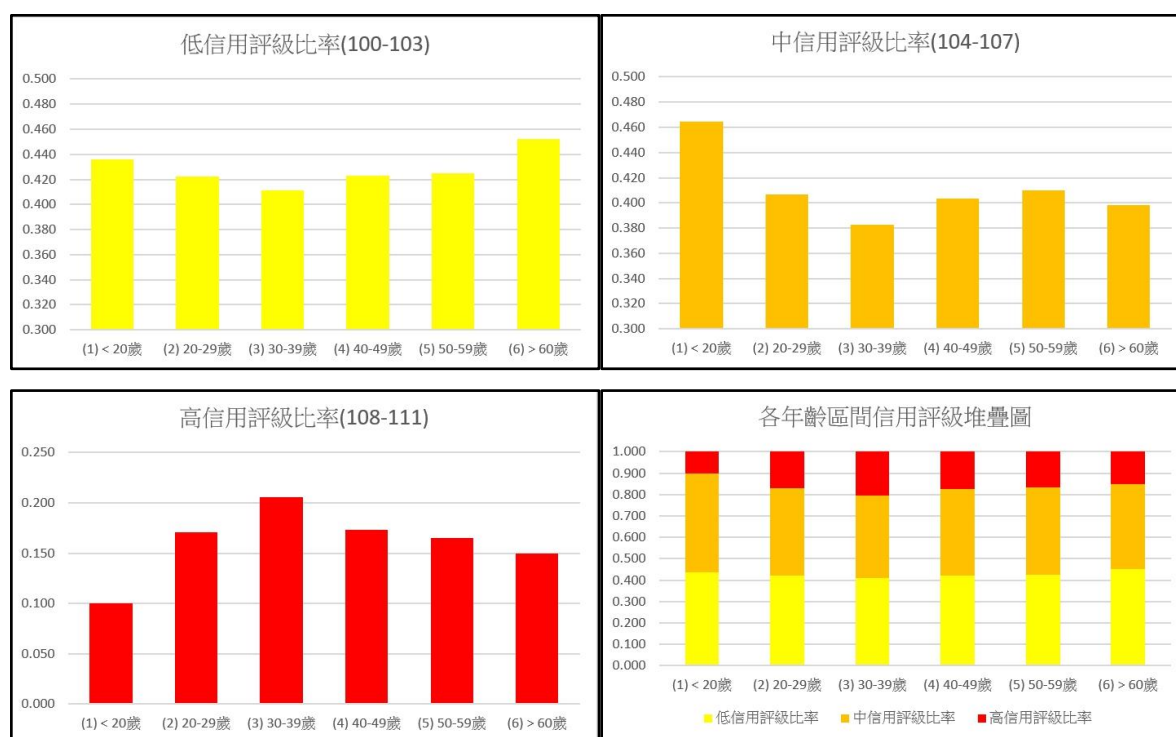
- 各年齡區間在一個月內的溢繳比率相差不大，差距在1%以內
- 小於20歲和大於60歲的年齡區間在三個月內的溢繳比率，略高於其他年齡區間
- 小於20歲的年齡區間在六個月的溢繳比率明顯高於其他年齡區間

### 分群基準：信用評級

以下為各年齡區間平均信用評級圖。可觀察到20歲以下平均信用評級略低於其他年齡區間，30-39歲則高於其他年齡區間。



以下幾張圖將信用評級拆成低(100-103)、中(104-107)、高(108-111)三個層級在各年齡層之間做比較



根據上面幾張圖，可得以下觀察：

- 60歲以上低信用評級比率略高過其他年齡區間
- 20歲以下中信用評級比率明顯高於其他年齡區間

- 30-39高信用評級比率超過2成，明顯高於其他；20歲以下高信用評級則明顯低於其他年齡區間
- 30-39歲的高信用評級比率高，低信用評級比率低，為信用評級優秀族群。

## 附錄：問題集

1、DCB交易資訊--購買失敗的定義以及可能因素為何？

2、DCB交易資訊--「成功平均購買金額佔信用等級比率(欄位名稱表，列216-218)」：欲釐清欄位分子/分母分別為何；尤分母為信用等級(100-111)，還是該信用等級的額度呢(且分別為何)？

3、繳款資訊--欠費(欄位名稱表，列14、19、24)欄位皆為N，意義為所有使用代收服務的用戶近期皆即時繳款、無欠費記錄嗎？若想得知繳款表現差的用戶資訊，會建議看哪個欄位的資料呢？