

2019 1.10



# 遊戲業營銷策略暨資料分析報告

GROUP4

組員：資管三 翁典(組長) 高從聞 林貞妮 統計四 陶培生

第一部分 前言與專案目標 .....	3
(一)    前言 .....	3
(二)    專案目標 .....	3
1. 協助優化現有公司營銷策略 .....	3
2. 提升營運績效.....	3
第二部分 資料概述&分析實作 .....	3
(一)資料概述 .....	3
1. 時間軸線 .....	3
日登入流量 .....	3
周登入流量 .....	4
月登入流量 .....	5
全玩家平均儲值金額.....	5
付費玩家平均儲值金額 .....	5
付費轉換率 .....	6
日流失率 .....	6
2. 個別帳號軸線 .....	7
儲值總次數 .....	7
儲值總金額 .....	8
總登入次數 .....	8
(二)分析實作 .....	9
1. RFM 模型 : .....	9
1.1 資料轉換成 R、F、M .....	8
1.2 將顧客分 8 類 .....	9
1.3 小結 .....	10
2. 登入天數預測模型 : .....	10
2.1 資料前處理 .....	10
2.2 建構複回歸模型 .....	11
2.3 模型調校與轉換 .....	11

	.....	11
2.4	調整後模型與小結 .....	11
	.....	12
3.	物品分析： .....	14
3.1	轉生次數與儲值.....	14
3.2	職業與物品購買 .....	17
3.3	RFM 結果與 Item Type .....	18
第三部分 行銷策略建議與結語 .....		20
(一)	行銷策略建議.....	20
(二)	結語 .....	21

## 第一部分 前言與專案目標

### (一)前言

我們將手上的資料藉由課堂上的知識輔以課後自學的技巧，先觀察並統計遊戲的流量，且計算出各項重要指標。決定我們的客戶群後，把儲值的玩家做 RFM 分類，並以複回歸分析預測玩家登入的天數，利用 RFM 和登入的結合，將玩家分類，以便能對各類用戶研發出有效、精準的行銷策略，協助客戶改善遊戲內容，增加收益。

### (二)專案目標

#### 1. 協助優化現有公司營銷策略

- 將顧客分類，鎖定所選的顧客
- 協助行銷活動的取捨點

#### 2. 提升營運績效

- 鎖定最有可能購買的有價值顧客
- 避免浪費行銷活動在購買可能性低的顧客身上

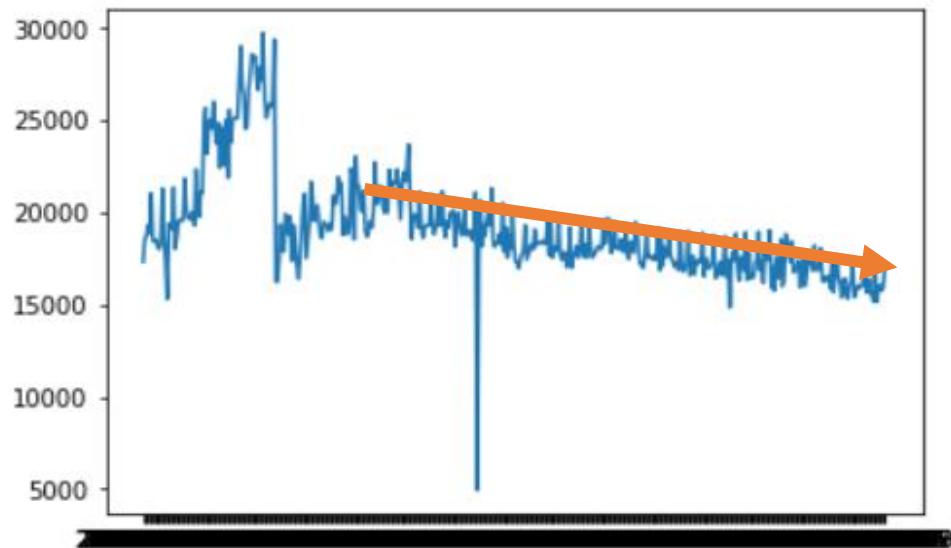
## 第二部分 資料概述&分析實作

### (一)資料概述

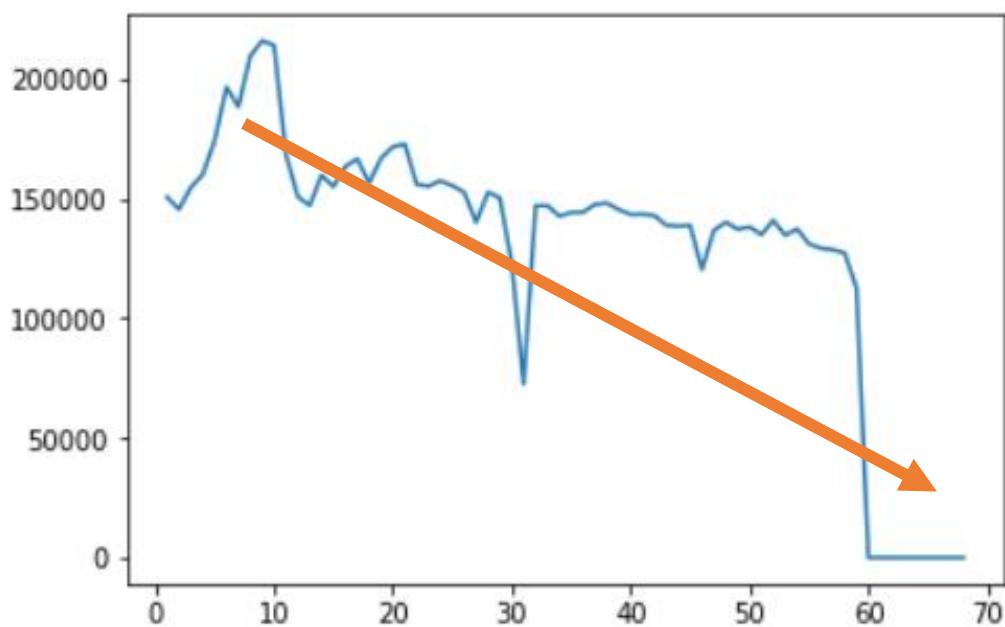
#### 1. 時間軸線

下方圖表呈現以時間軸線區隔而成的資料概況

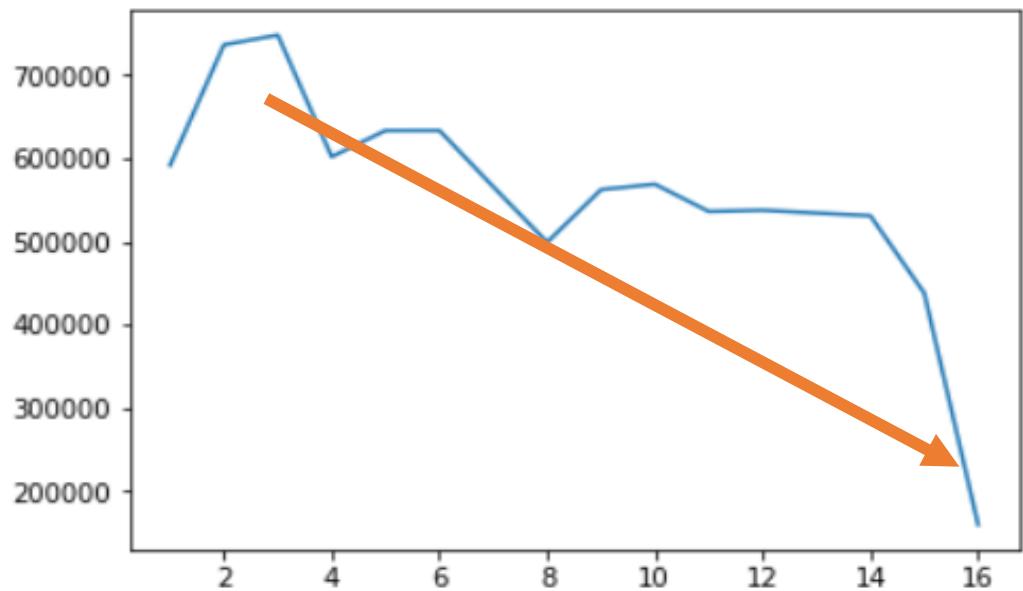
- 日登入流量



● 周登入流量

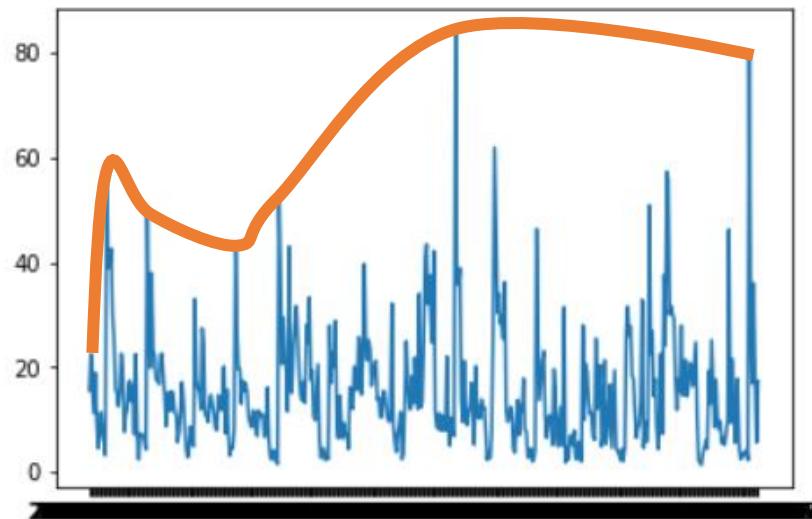


- 月登入流量

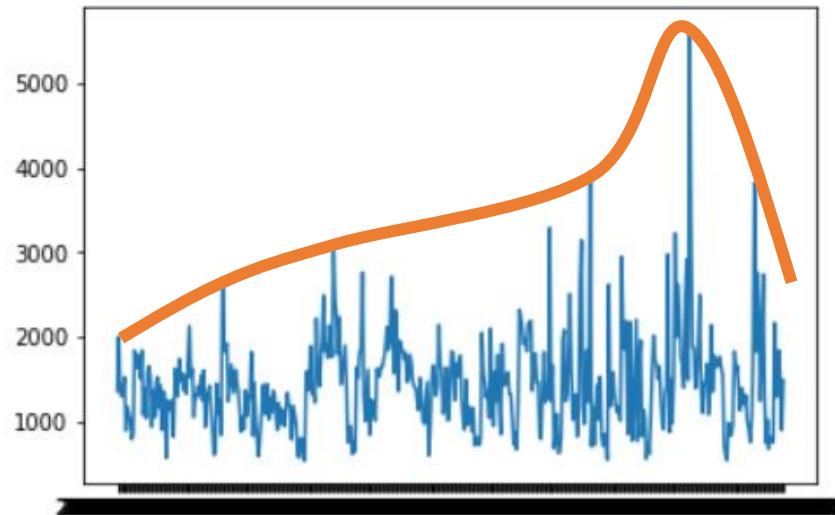


◎從日/周/月三個不同時長的折線圖來看，登入流量有逐漸下滑的趨勢。

- 全玩家平均儲值金額(ARPU)

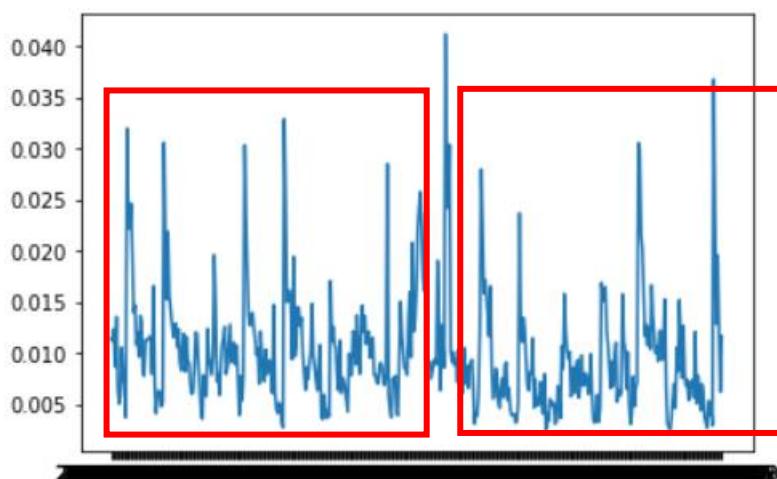


- 付費玩家平均儲值金額(ARPPU)



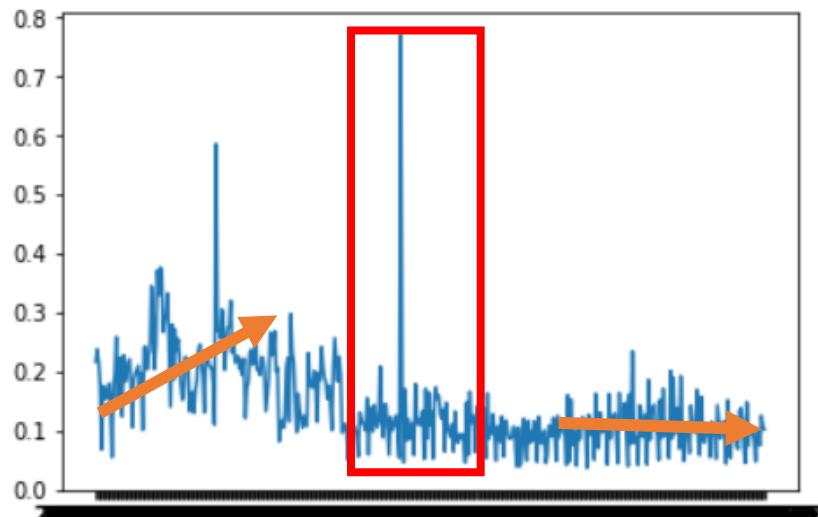
◎ARPU 與 ARPPU 的分布來看儲值玩家有持續儲值的情況且到後期之前高峯值有增加趨勢，後期以降才明顯下降。

- 付費轉換率(PUR)



◎從每日儲值玩家佔全玩家的比例來看，數據略為呈現左右對稱的 pattern ，波動狀況有點循環性。

- 日流失率(Churn rate)



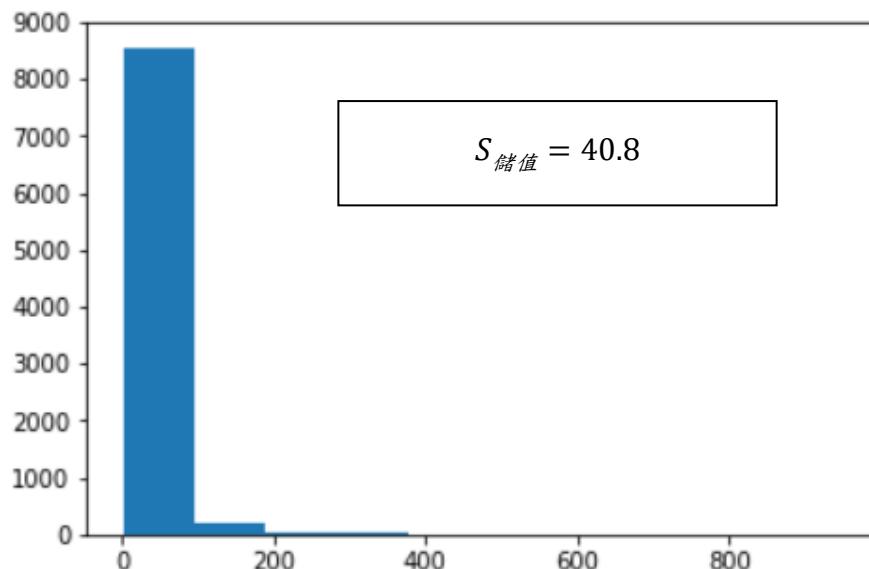
◎流失的情況在初期呈現上升趨勢，中間部分有突然的猛升，後期漸趨平緩。

小結：綜觀以上圖表，玩家隨著時間流失的幅度應當大過儲值玩家增加與儲值玩家增加其儲值金額的幅度，導致 ARPU & ARPPU 有峰峰高的趨勢。

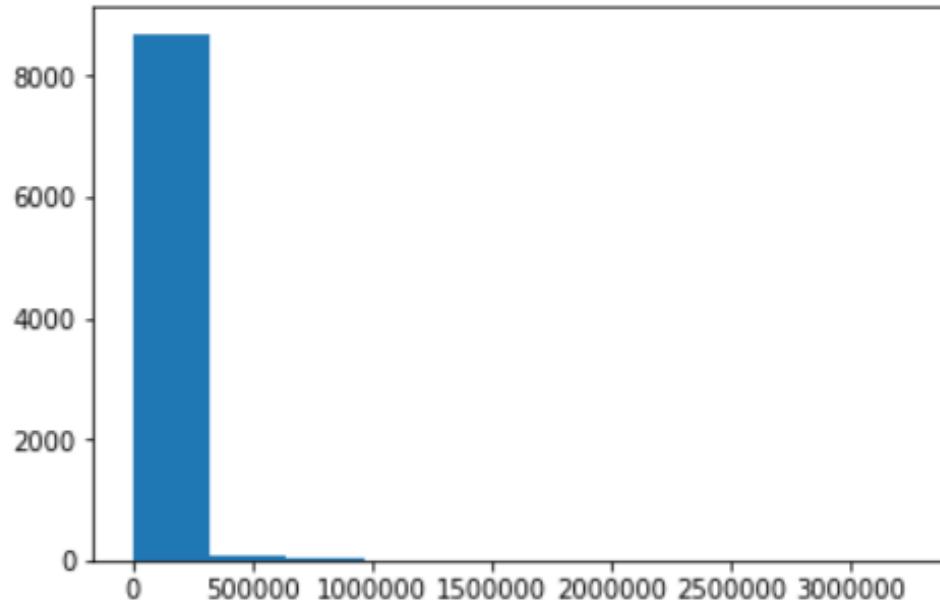
## 2. 個別帳號軸線

下方圖表呈現以個別帳號軸線區隔而成的資料概況

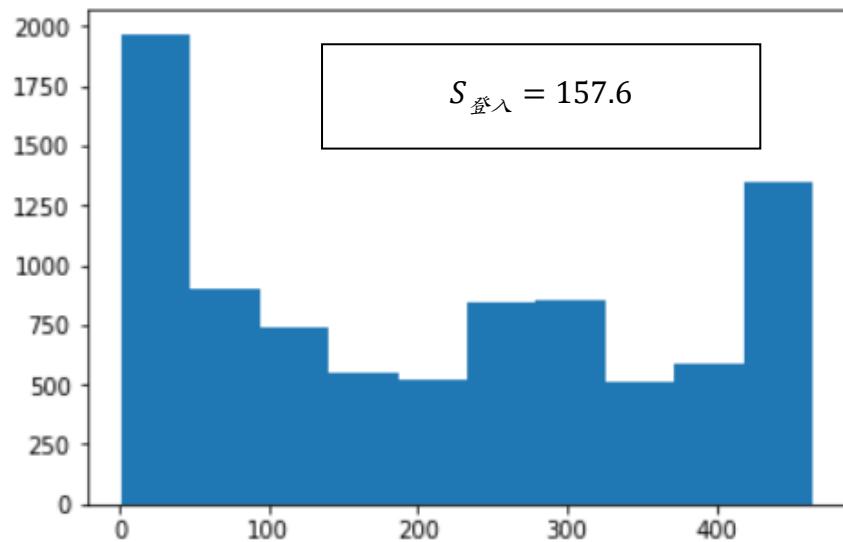
- 儲值總次數



- 儲值總金額



- 總登入次數



◎綜觀以上三圖絕大多數的玩家儲值意願並不高，儲值離散的情況遠小於登入次數離散的狀況，顯示儲值次數普遍低落。

## (二) 分析實作

### 1. RFM 模型：

#### 1.1 資料轉換 R、F、M

值資料轉換為 R、F、M 三個指標，分別是 Recency, Frequency, Monetary。Recency 最後一次儲值時間離現在有多少天，Frequency 儲值頻率，Monetary 儲值總金額。

	儲值遊戲帳號	交易成功時間	儲值金額	recency	frequency	monetary_value	儲值遊戲帳號
0	3352496	2017-01-31	500				1289872 379 days 38 9600
1	3352496	2018-02-19	500				1290144 100 days 63 105480
2	3458960	2017-07-20	500				1292240 729 days 1 150
3	3458960	2017-08-03	1000				1293472 519 days 1 150
4	3458960	2017-08-25	150				1317520 422 days 1 50

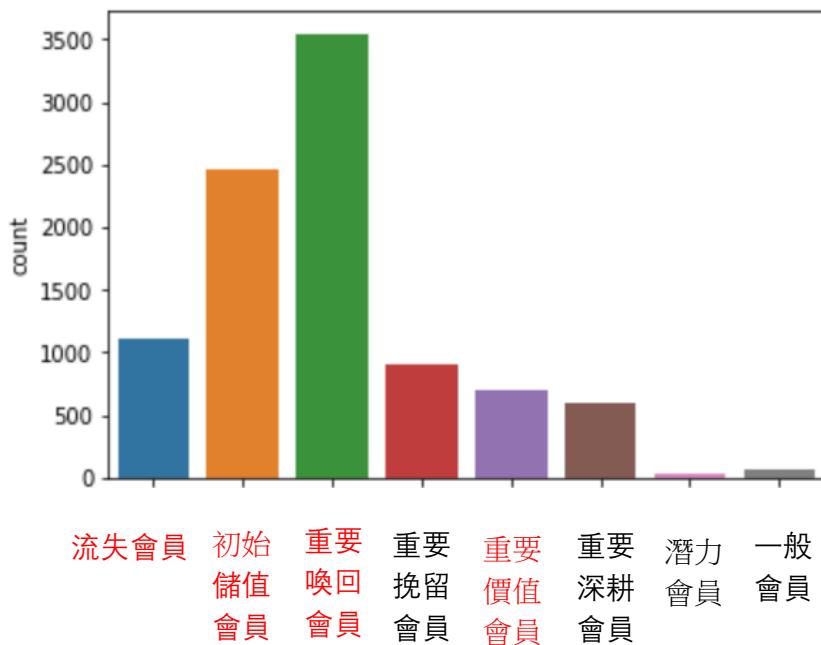
  

	儲值遊戲帳號	recency	frequency	monetary_value	R_Quartile	F_Quartile	M_Quartile	RFMClass	Total Score
	1289872	379.0	38	9600	3	1	2	312	6
	1290144	100.0	63	105480	1	1	1	111	3
	1292240	729.0	1	150	5	5	5	555	15
	1293472	519.0	1	150	4	5	5	455	14
	1317520	422.0	1	50	3	5	5	355	13

#### 1.2 R、F、M 有兩個方向高、低，因此將顧客分成 8 個類別

儲值遊戲帳號	R_Quartile	F_Quartile	M_Quartile	RFMClass	分類
1289872	3	1	2	312	8 流失會員
1290144	1	1	1	111	4 近期儲值會員
1292240	5	5	5	555	5 重要喚回會員

1383600	2	2	3	223	3 重要深耕會員
45484330	3	4	2	342	6 一般會員
1341488	2	4	4	244	1 重要價值會員
1505952	2	5	2	252	2 潛力會員
1330672	5	3	3	533	7 重要挽留會員



### 1.3 小結

利用 RFM 將玩家分為 8 個不同的類別，接下來會再結合以下的預測

登入天數，進行玩家更細項的分類，以利針對玩家做精準行銷。

而上面標註紅字的將是我們在以下結合預測登入天數後將會著重行銷的顧客類別。

## 2. 登入天數預測模型：

### 2.1 資料前處理

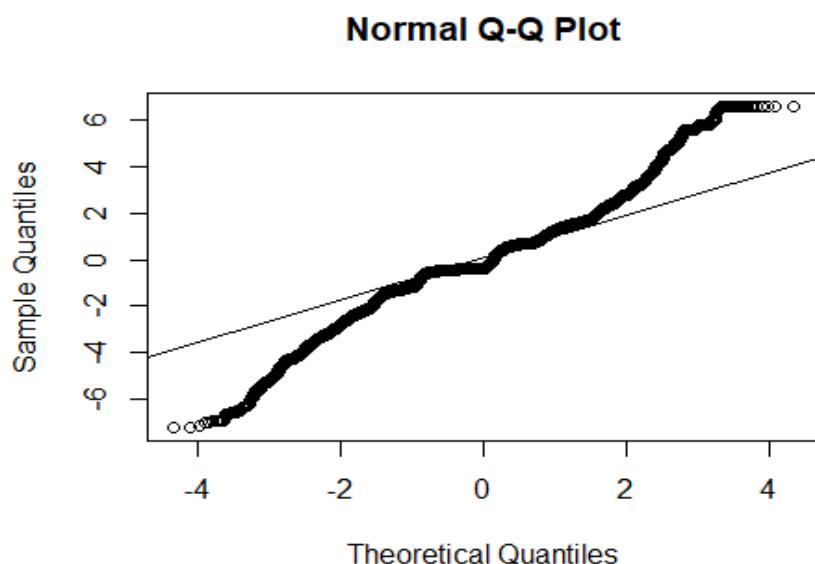
將原始資料讀入 R 後更改 data type 使之容易操作並以時間為軸切割資料，透過迴圈與條件控制式計算出個別帳號上週與上個月的登入次數。

## 2.2 建構複回歸模型

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ ，其中  $Y$ (應變數)表示某玩家本周登入天數預測值， $X_1$ 、 $X_2$ (自變數)分別表示上週登入天數與上個月登入天數。

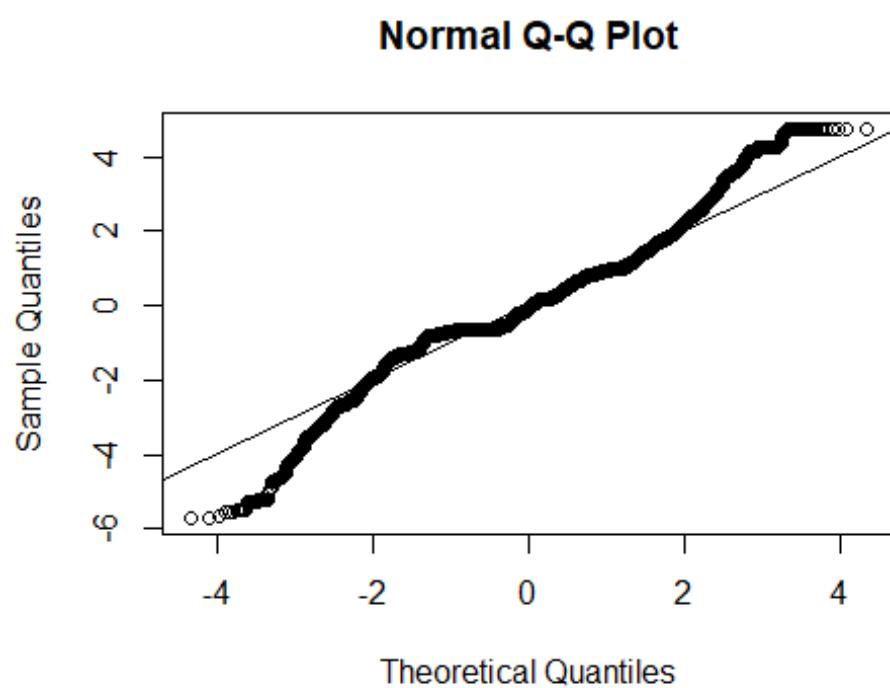
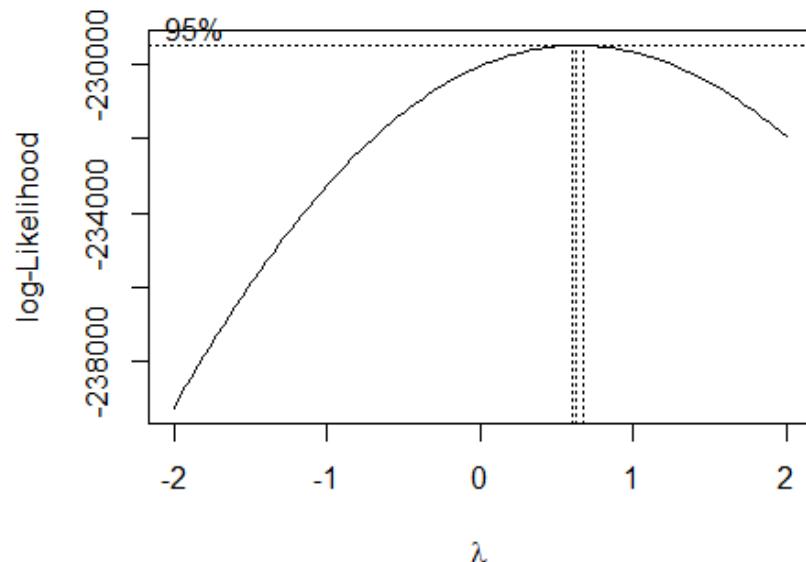
## 2.3 模型調校與轉換

從以下 Q-Q Plot 得知變數的分佈是雙峰，不是常態分配，造成 residuals 偏離嚴重，因此使用 Boxcox Transformation 調整。



## 2.4 調整後模型與小結

經推算得知  $\lambda = 0.62$  時 Boxcox 效果最為顯著，將  $\lambda = 0.62$  帶入原複回歸模型可發現 R-square 有所提升和 residuals 之分布也相較平緩。  
(見下圖)



小結：此複迴歸之 R-sqaure = 0.78 代表預測項(下週登錄天數)可被解釋的部分高達 78%兩自變數獨立之 P value 和整體之 P value = 2.2e-16 也非常顯著。(見下圖)

```

Call: +
lm(formula = (nextweek^(0.62) - 1)/0.62 ~ lastweek + lastmonth, +
  data = sumtest) +
+
Residuals: +
  Min    1Q Median    3Q   Max +
-5.7073 -0.6529 -0.0961  0.6938  4.7482 +
+
Coefficients: +
            Estimate Std. Error t value Pr(>|t|) +
(Intercept) -1.599e+00 7.633e-03 -209.4 <2e-16 *** +
lastweek      6.159e-01 1.761e-03   349.8 <2e-16 *** +
lastmonth     1.136e-02 5.652e-05   201.1 <2e-16 *** +
--- +
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 +
+
Residual standard error: 0.999 on 70576 degrees of freedom +
Multiple R-squared:  0.7805, Adjusted R-squared:  0.7805 +
F-statistic: 1.254e+05 on 2 and 70576 DF, p-value: < 2.2e-16 +

```

透過預測玩家下週登入天數，我們可以結合 RFM 將玩家分為以下 24 種不同類型，但最後我們會針對以下紅色框框的 4 種玩家進行精準行銷。

	重要價值	重要喚回會員	重要挽留會員	重要深耕會員	潛力會員	初始會員	一般會員	低消費會員
上線多	2.311215252	5.314730003	1.693471083	2.790499521	0.11715838	15.72052402	0.117158377	2.71594419
上線一般	0.660347215	1.214186814	0.468633507	0.62839493	0.04260305	3.653211205	0.106507615	0.745553307
上線少	4.38811375	31.24933433	7.455533071	2.843753328	0.09585685	6.848439663	0.4153797	8.403450847

會針對這 4 種玩家是因為

上線多的重要價值也就是 VIP，是公司勢必維持的客戶來源。

上線少的重要喚回會員，是第一比例高的，因此我們必須做出對這類玩家做出方案，上線少或許是對遊戲感到疲乏，或是他們無法從遊戲中獲得成就感，因此我們必須可以藉由增加遊戲豐富度、改版、

玩家見面會等等來改善。

上線多的初始會員，比例是第二高。

上線多的低消費會員，上線多代表他們是熱愛遊戲的。

### 3 物品分析：

- 分析目標：透過物品的分類與區隔希望能將不同的物品精準行銷給淺在的客戶。結合儲值資料找出遊玩時間長短與儲值金額的關係。
- 資料觀察與處理：將給定各項資料表輸入 PowerBI 建立關聯式資料庫，並合併部分報表。
- 分析：

#### 3.1 轉生次數與儲值

每個帳號 id 的儲值次數與儲值總金額建立資料表，並分為是否有轉生級別兩個面向，選擇以轉生次數分類是衡量一個玩家進行遊戲的時間長短，由於不能直接由 character 資料以及登入次數中準確衡量遊玩時數，因此由轉生次數做分類分別建表如圖。

這裡以轉生次數=0 但有儲值的資料表舉例

```
In [5]: df_reborn0.head()
```

Out[5]:

	<b>id</b>	<b>count</b>	<b>amount</b>
<b>0</b>	1292240	1	150
<b>1</b>	1383568	2	200
<b>2</b>	1386608	2	600
<b>3</b>	1409520	1	150
<b>4</b>	1434944	2	300

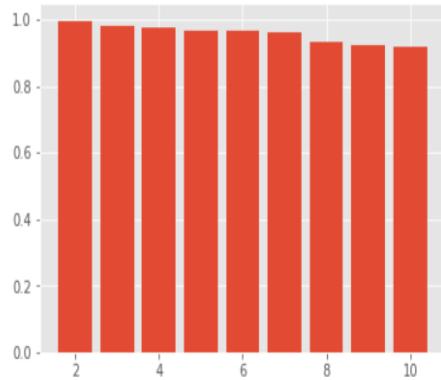
以 python 的 pandas 插件讀入 dataframe 然後做 k-means 分群

```
In [10]: X = df_reborn0[['count','amount']].values
```

判斷 k-means 分群績效：

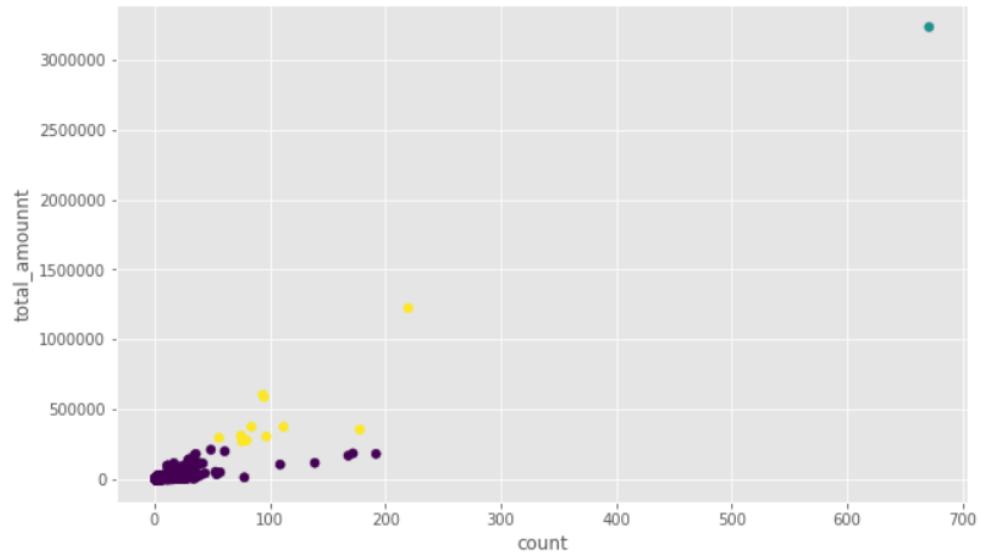
```
In [11]: # 週圖
silhouette_avgs = []
ks = range(2, 11)
for k in ks:
    kmeans_fit = cluster.KMeans(n_clusters = k).fit(X)
    cluster_labels = kmeans_fit.labels_
    silhouette_avg = metrics.silhouette_score(X, cluster_labels)
    silhouette_avgs.append(silhouette_avg)

# 作圖並印出 k = 2 到 10 的績效
plt.bar(ks, silhouette_avgs)
plt.show()
print(silhouette_avgs)
```



[0.9964322750740837, 0.9825494862912768, 0.9780726581345067, 0.9677801484477176, 0.9685815807895185, 0.9626123701074827, 0.931216290351197, 0.9243606425177254, 0.9190644257263894]

```
In [12]: km = KMeans(n_clusters=3) #K=2群
y_pred = km.fit_predict(X)
plt.figure(figsize=(10, 6))
plt.xlabel('count')
plt.ylabel('total_amount')
plt.scatter(X[:, 0], X[:, 1], c=y_pred) #C是第三維度 已顏色做維度
plt.show()
km.cluster_centers_ #各群中心點(X, Y)的位置
```



Out[12]: array([[3.10704337e+00, 3.19537565e+03],
 [6.71000000e+02, 3.23488000e+06],
 [1.03583333e+02, 4.39104167e+05]])

由於績效評估過於接近，無法得出顯著結論。

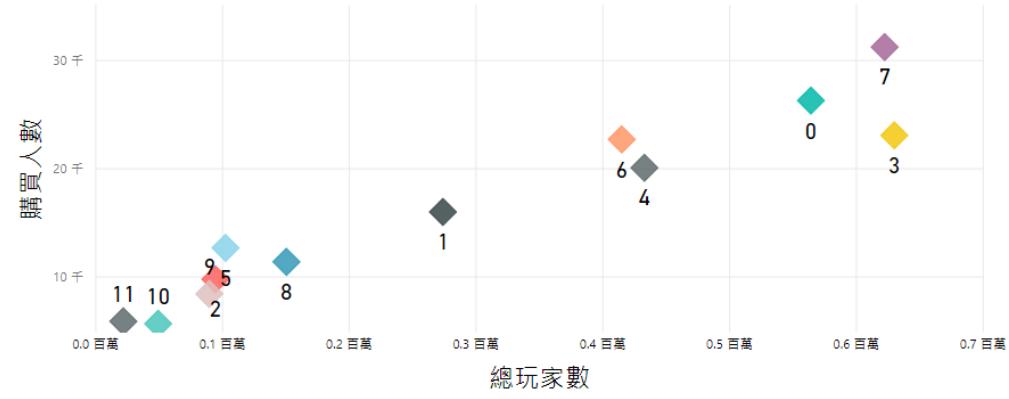
(其他 k-means 分析結果放於 5. 分析結果)

### 3.2 職業與物品購買

本資料共有 11 種職業(1~11)觀察資料中各職業購買物品的種類，得知物品沒有職業分別。分別統計各職業總人數與各職業購買人數，並計算出購買比例。

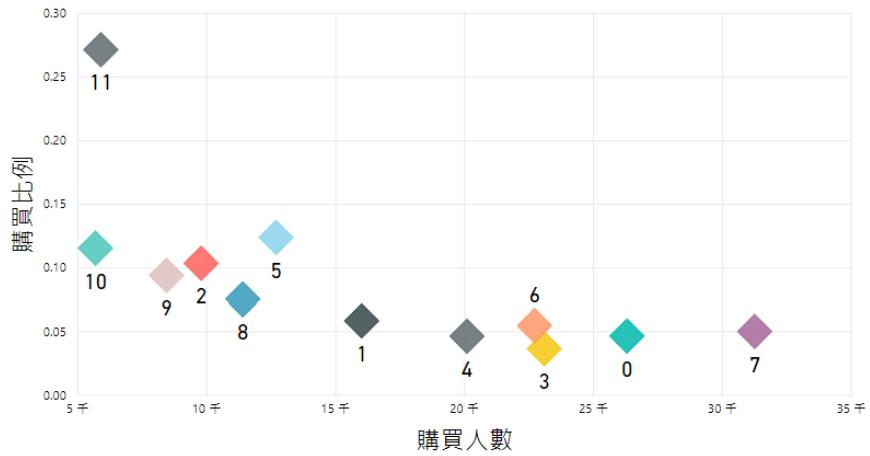
職業總人數與購買人數分布圖

職業 ◆ 0 ♦ 1 ♦ 2 ♦ 3 ♦ 4 ♦ 5 ♦ 6 ♦ 7 ♦ 8 ♦ 9 ♦ 10 ♦ 11



職業購買比例與人數分布圖

職業 ◆ 0 ♦ 1 ♦ 2 ♦ 3 ♦ 4 ♦ 5 ♦ 6 ♦ 7 ♦ 8 ♦ 9 ♦ 10 ♦ 11



11 號職業有最少的遊玩人數，卻有最高的購買比例(27%)

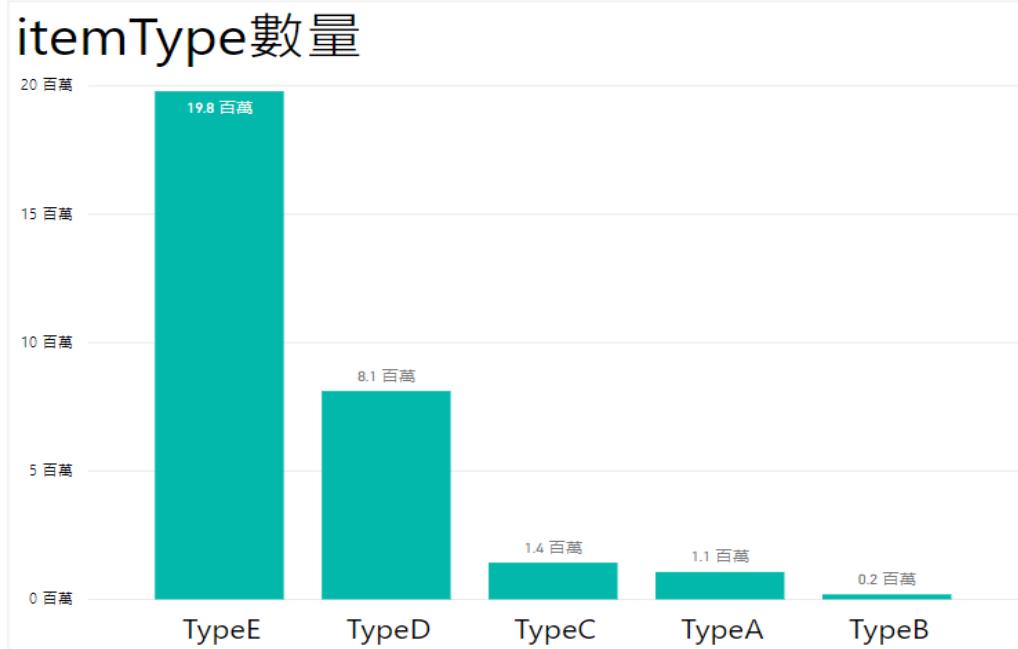
3 號職業有最多的遊玩人數，但僅有 5%購買比

推測職業之間不平衡，例如：職業 0, 3, 7 可能過強。或是職業對

新手不友善導致需要較多道具輔助，例如：職業 10, 11。

### 3.3 RFM 結果與 Item Type

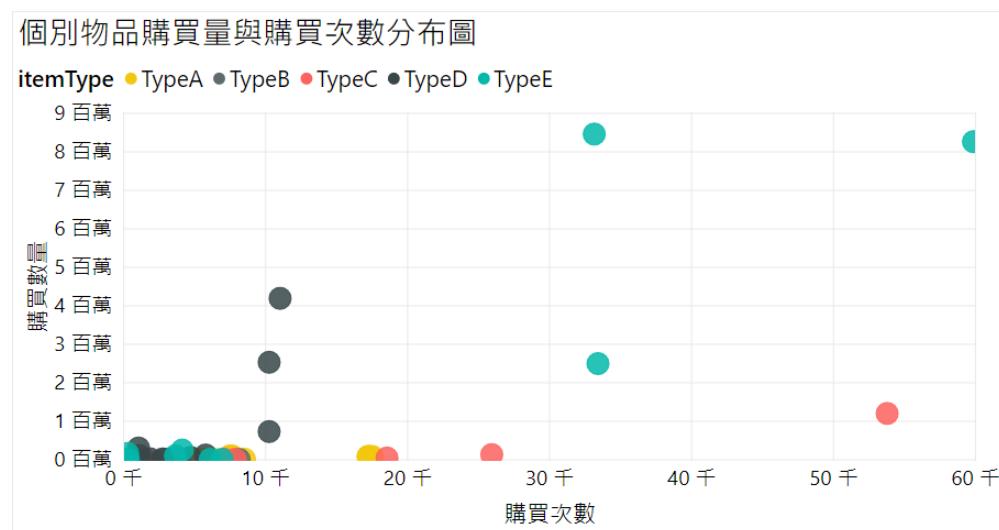
由資料觀察已知：Item Type 分為 A, B, C, D, E 五種，所有 157 種



貴重物品中，所有物品皆有所屬的 Type。

銷售量：

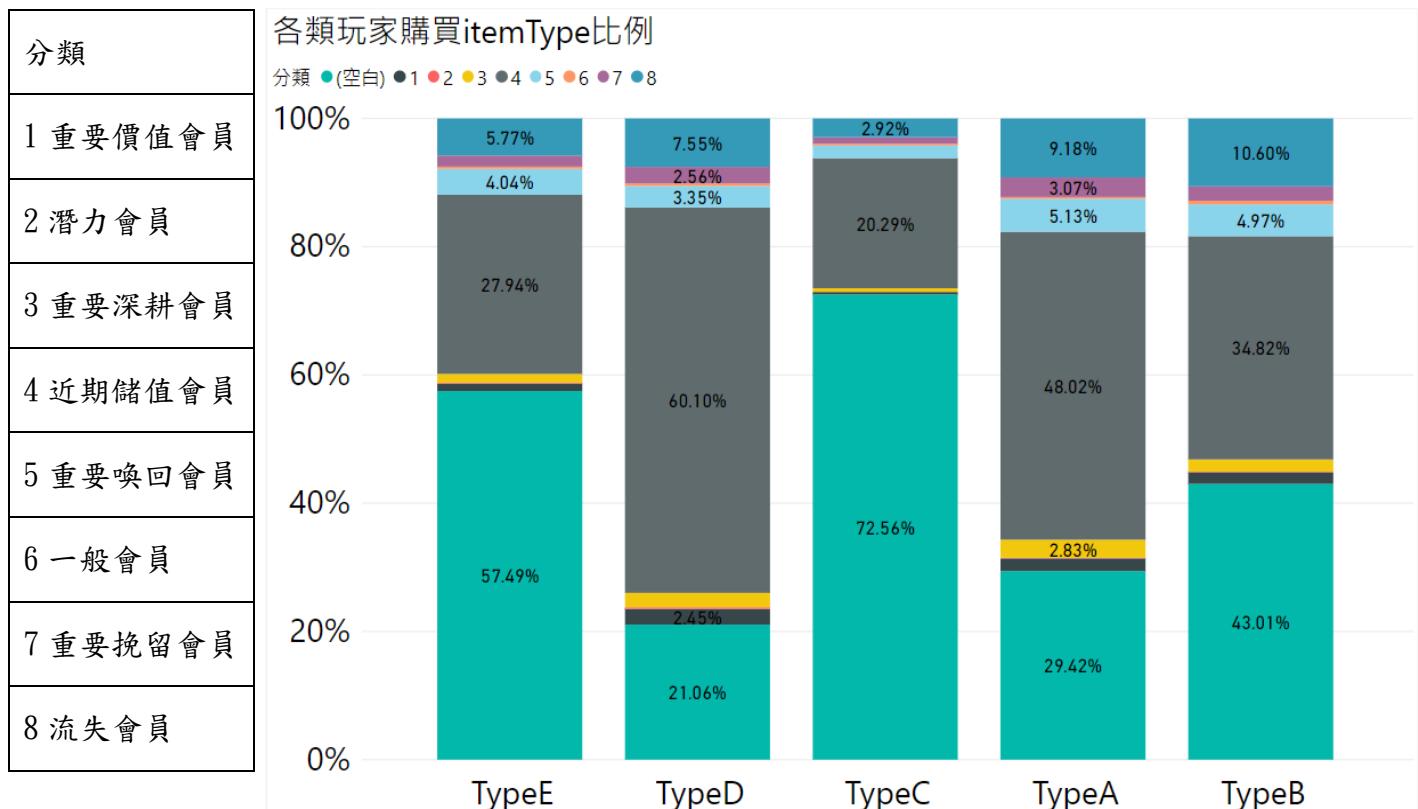
TypeE(64.67%)>TypeD(26.51%)>TypeC(4.68%)>TypeA(3.49%)>TypeB(0.64%)



觀察物品被消費的次數和購買量作出以下推測：

TypeE、TypeD 因為消費次數和消費量高屬於消耗品，而 TypeD

又是層級更高的種類。TypeA，TypeC 購買數量不多屬於耐用品，TypeB 則屬於具有套裝效果的套裝物品。將 RFM 結論中的 1~8 類儲值玩家建表，與不儲值玩家作累積長條圖，分析各類玩家購買傾向。(下圖各類玩家圖例可對照左邊)



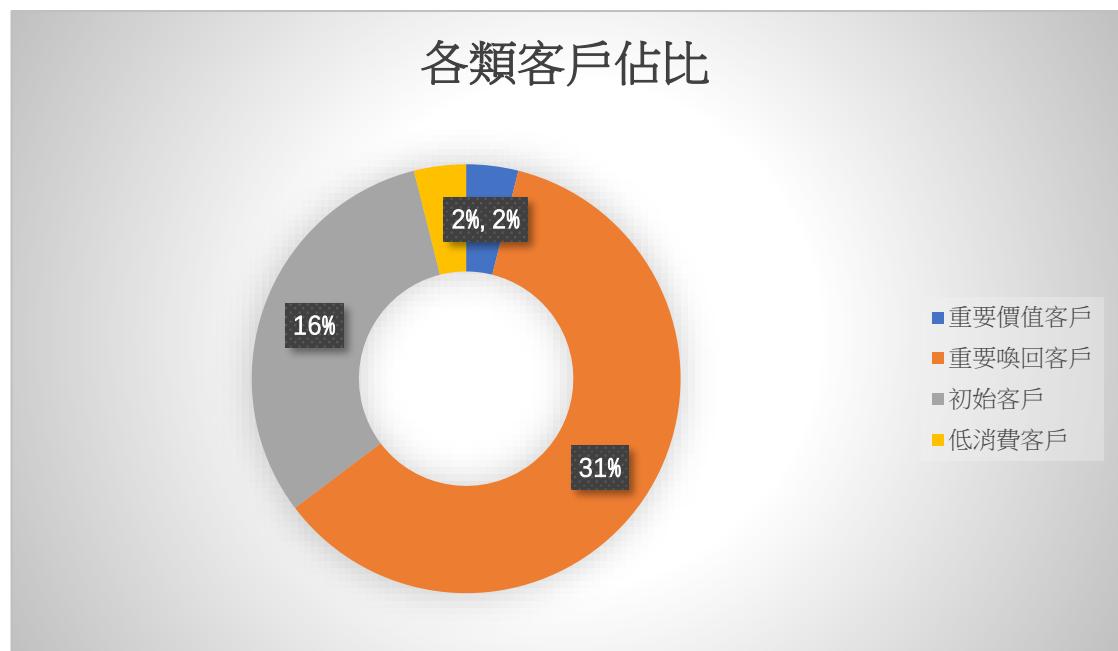
(註:圖中標示”空白”的部分是指沒有儲值的玩家。)

可以得知，付費儲值玩家佔 TypeD 消費大宗，TypeD 應為代價高昂的消耗品，非儲值玩家不易取得。TypeE 應為遊戲中的必需品。TypeC 被非儲值玩家喜愛，可能是武器或防禦道具。TypeA 可能也是武器或防禦道具，但購買優先度不如 TypeC、TypeB 購買量極低，付費、非付費用戶購買比例接近，可知套裝物品不構成對玩家的吸引力。

## 第三部分 行銷策略建議與結語

### (一) 行銷策略建議

從各類玩家中，以客戶價值和所佔比重選擇其中四類：重要價值客戶、重要喚回客戶、初始客戶、低消費客戶。



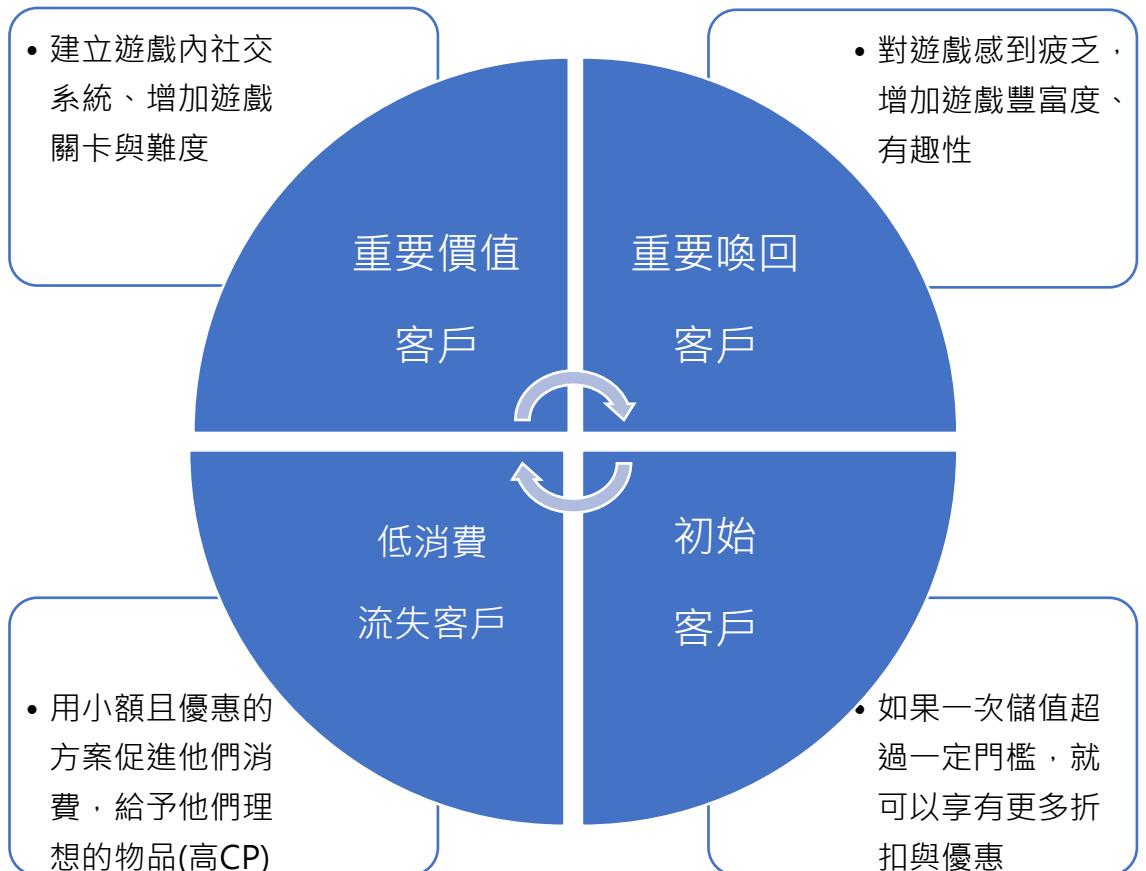
定義：

重要價值客戶：上線多消費的VIP。

重要喚回客戶：對遊戲感到疲乏的高消費玩家。

初始客戶：新進消費的玩家，消費頻率與消費量不高。

低消費客戶：上線多，熱愛遊戲但消費預算有限的玩家。



## (二)結語

經由一連串的分析與實作我們重新認識大數據，並了解到實務界對於大數據的應用以及它的潛力，大數據分析可以幫助公司找到不易發現的問題，也能從中開發新的客戶，在解決問題的同時替公司尋求更多收入。

經由上手的實作經驗，我們也學習如何應用不同的程式技巧去達成所求，有了這次處理巨量資料的寶貴經驗，以後有相關問題也能更快進入狀況。