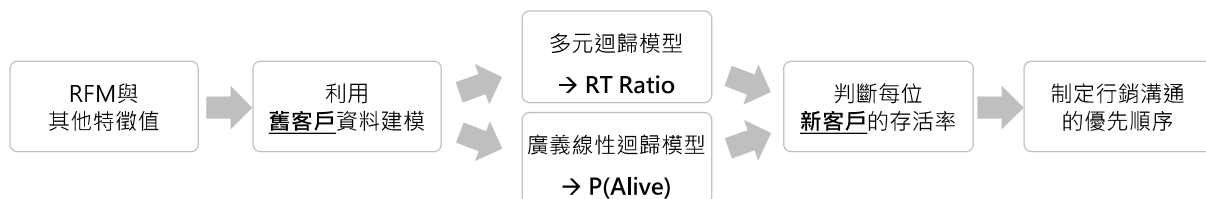


## Final Project - AsiaYo

### 一、研究目的與脈絡

原題目為：Customers Retention 是品牌的一個重要指標，有效的瞭解公司用戶並且利用 RFM 做分類並對其進行數據分析，並制定營銷策略。其透過 RFM 的分類找出在一定時間區隔內（如：3 個月），各群的回購人數占比，並依此判斷各群的存活率(Retention Rate)，最終以利 AsiaYo 公司決定在給定的行銷預算下，要先針對哪一群做行銷溝通。然而，本組認為每位用戶的訂房週期不同，例如：某 A 固定每年出遊一次，因此對某 A 來說，其訂房週期為一年，但若統一以 3 個月判斷之，某 A 會被列入不存活客群的分類中，但實際上只是某 A 的訂房週期尚未到。因此本組認為用統一時間區隔決定的回購人數佔比，是無法用來判定各群的存活率。

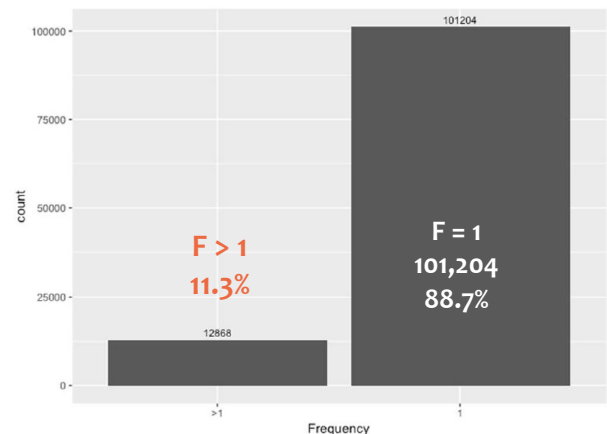
所以本組改以利用舊客戶的資料的 RFM 與更多的特徵值（如：客戶偏好訂房類型、客戶偏好用何種平台下訂等）去建立**(1)多元迴歸模型**去求出新客戶的 RT Ratio 與**(2)廣義線性迴歸模型**去求出新客戶的 P(Alive)，最後**判斷新客戶的存活率**，並將 RT Ratio 與 P(Alive)設定為兩個維度去做新客戶的分群，依照這兩個維度去制定行銷溝通的優先順序，整體脈絡如下圖。



其中，在此定義的新客戶與舊客戶分別為在資料集中購買頻率為乙次與大於乙次者，其分別占比呈現於下表。

命名	占比	定義
新客戶	88.7%	資料區間內購買次數=1 者
舊客戶	11.3%	資料區間內購買次數>1 者

而在多元迴歸模型中，本組欲求出的 RT Ratio 為 Recency/訂房週期 T，其意義是若  $RT\ Ratio \leq 1$  則認為客戶回購機率較高，若 RT Ratio 越大則認為客戶可能想離開或者已經離開 AsiaYo 平台。



### 二、資料前處理 - 數值變數

為了針對 AsiaYo 用戶進行 K-prototypes 分群與 Multiple Linear Regression (MLR)，前置工作為將 AsiaYo 提供的「one row per order id」訂單資料彙整成「one row per user id」的資料型態，由於一位用戶可能有多筆訂單資料，因此部分欄位會以機率的方式表示（e.g. 旅遊地點、旅宿房型等），以便後續化為類別變數。資料前處理欄位與邏輯說明如下：

#### 1. “TtlContri”

該用戶在訂單資料中所有台幣金額(“*twd\_amount*”)加總。

#### 2. “Recency”

訂單資料最晚日期為 2019-11-09，因此設定 2019-11-10 為 tag date，計算該用戶最後一筆訂單日期距離 tag date 的天數。

#### 3. “TMed”

計算該用戶每筆訂單之間間隔的天數(訂購週期 T)，再取其中位數為“TMed”。

## 4. “RTRatio”

該用戶“Recency”除以“TMed”的比值。

## 5. “Frequency”

本組認為用戶在一次旅行中訂幾次房其實跟旅遊地點、天數、人員組成等因素有關，因此在分析用戶 retention rate 時，「旅行的頻率」比「訂房的頻率」有意義。而本組假設用戶會在同一天將一次旅行所需的所有訂房完成，因此本欄計算的是該用戶在訂單資料中所有不重複日期數量。

## 6. “Monetary”

該用戶平均每次旅行的花費，計算方式為“TtlContri”除以“Frequency”的比值。

## 7. “AvgPmt”

該用戶平均每人每晚的花費，計算方式為“TtlContri”除以“TtlGuest”（總訂單 Guest 數）與“TtlNight”（總訂單 Night 數）的比值。

## 8. “TtlQpon”

該用戶總共在訂房時使用 Coupon 次數。

## 9. “DailyQpon”

該用戶平均每次旅行在訂房時使用 Coupon 次數，計算方式為“TtlQpon”除以“Frequency”的比值。

## 10. “TtlNote”

該用戶總共在訂房時撰寫文字附註次數。

## 11. “DailyNote”

該用戶平均每次旅行在訂房時撰寫文字附註次數，計算方式為“TtlNote”除以“Frequency”的比值。

## 12. “Country”

原訂單資料既有欄位，為該用戶國籍。

## 13. “Locale”

原訂單資料既有欄位，為該用戶網站語系版本。

## 14. Book status

訂單資料中訂房狀態(“book\_status”)共有四種：Accepted, Rejected, Cancelled, Unconfirmed，本組將排除 Unconfirmed 後的訂單定義為「有效訂單」(“EffBkOrder”)。本組先分別加總該用戶所有歷史訂單中狀態為 Accepted, Rejected, Cancelled 的訂單數量，考量到各用戶訂單數量不同，直接以訂單總數做為比較基準有失公允，因此將各用戶三種狀態訂單數分別除以該用戶“EffBkOrder”數化為機率(“PctAcc”, “PctRej”, “PctCan”)表示。以某用戶舉例如下：

Accepted	Rejected	Cancelled	Unconfirmed	EffBkOrder	PctAcc	PctRej	PctCan
5	3	2	1	10	0.5	0.3	0.2

## 15. Others

針對各用戶訂單狀態為 **Accepted** 的訂單，本組進一步探討各項特徵，同樣考量到各用戶訂單數量不同，直接以該特徵出現次數直接加總做為比較基準有失公允，因此將各用戶該特徵次數加總後分別除以該用戶“Accepted”訂單數化為機率表示。以某用戶「下單平台」(“platform”)舉例如下：

Accepted	Android	iOS	Mobile	PC	Android (Pct)	iOS (Pct)	Mobile (Pct)	PC (Pct)
5	2	0	1	2	0.4	0	0.2	0.4

以同樣邏輯計算整理的特徵包括：人數(Guests ; PctGuest1, PctGuest2, ..., PctGuest45)、小孩人數(Child ; PctChild0, PctChild1, ..., PctChild13)、房間數(Rooms ; PctRm1, PctRm2, ..., PctRm24)、過夜數(Nights ; Night1, Night2, ..., Night49)、下單平台(Platform ; Android, iOS, Mobile, PC)、旅宿國家(Bnb Country ; PctHK, PctKR, ..., PctTH)、旅宿類型(Bnb Type ; PctBnbTypAPT, PctBnbTypEHot, ..., PctBnbTypVilla)、星期幾下單(Order Date ; PctOrdMon, PctOrdTue, ..., PctOrdSun)、下單時段(Order Time ; PctOrdMorning, PctOrdAfternoon, ..., PctOrdMidnight)、入住月份(Check-in Date ; PctCheckinJan, PctCheckinFeb, ..., PctCheckinDec)、提前幾天下單(Reservation ; Order Date 與 Check-in Date 間隔天數 ; PctOrd3-, PctOrd4+, ..., PctOrd90+)。

### 三、資料前處理 - 類別變數

#### 1. 類別變數邏輯

为了更好的了解每個變數是否能夠顯示出顧客不同的習慣或偏好，本組將前部分所得到的數值變數與機率轉為類別變數，共新增十個類別變數如下：

Variables	Description	Variables	Description
<b>NightType</b>	用戶訂房偏好停留幾晚	<b>DoW</b>	用戶偏好的訂房時間
<b>ChildType</b>	用戶是否有帶小孩出遊	<b>OrderPre</b>	用戶偏好在入住時間多少天前訂房
<b>GuestNum</b>	用戶偏好出遊的團體大小	<b>Season</b>	用戶出遊季節偏好
<b>Device</b>	用戶偏好使用何種管道在該平台訂房	<b>Area</b>	用戶「出發地」所處的區域
<b>BnbType</b>	用戶偏好訂購何種房型	<b>Travel_dist</b>	用戶偏好的旅途距離長短

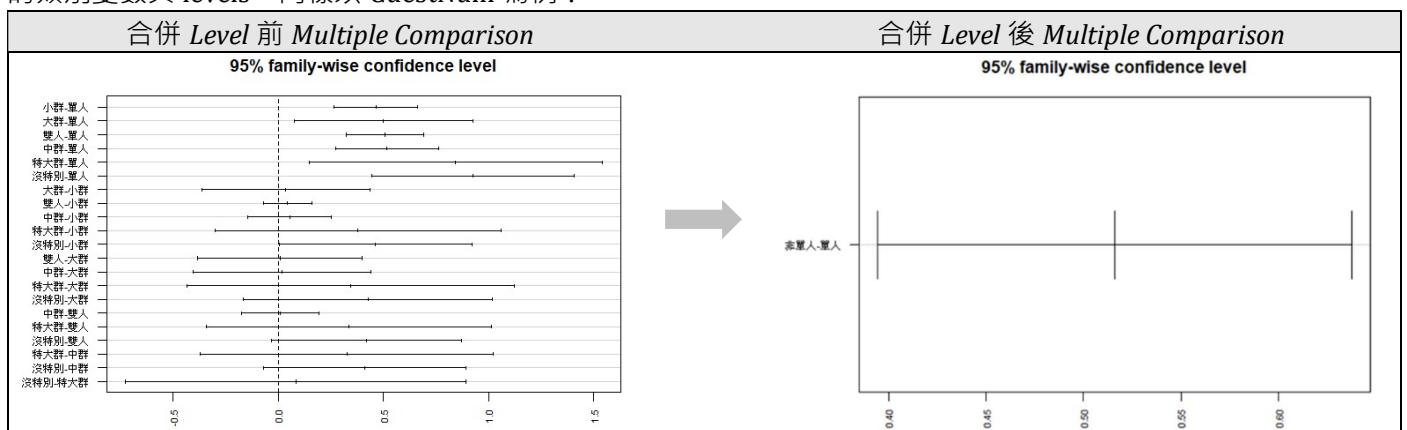
各個類別變數的命名邏輯與切分 level 的邏輯相似，取「大於一定比例(50%)」為一個 level。目的是希望看出顧客有沒有特別的偏好或習慣。以 **GuestNum** 為例：

Levels of "GuestNum"	Rules
<b>single</b>	PctGuest1(單人訂房)大於等於 50%
<b>nonsingle</b>	other

另外較為特別的類別變數為 **Travel\_dist**，定義方式首先將顧客的出發國家分為 EastAsia、EastSouthAsia、以及 Other 三個區域，再以不同出發與到達區域來判斷旅途的長短，分別為 **P\_Local**(當地旅遊)、**P\_Short**(短程旅遊)、**P\_Far**(長程旅遊)、以及其他等，藉以了解用戶旅遊長短偏好(其他類別變數請詳 八、附錄)：

Levels of "Travel_dist"	Rules
<b>P_Local</b>	出發地和目的地相同的比例大於 50%
<b>P_Short</b>	出發地和目的地為短程 ( EastAsia 到 EastAsia 或 EastSouthAsia 到 EastSouthAsia ) 的比例大於 50%
<b>P_Far</b>	出發地和目的地為中程 ( EastAsia 到 EastSouthAsia 或 EastSouthAsia 到 EastAsia ) 的比例大於 50%，以及出發地和目的地為長程 ( Others 到 EastSouthAsia 或 EastAsia ) 的比例大於 50%
<b>No Tendency</b>	其他不包含以上分類者
<b>N_Local</b>	不包含在 P_Local 者

過程中，由於最初本組切類別變數的 levels 時，是以想法與一般直覺，有切分過細之疑慮，因此本組將各個變數做 Multiple Comparison 了解 levels 間是否有顯著差異，並思考若沒差異而選擇合併是否會不合邏輯，得到最終的類別變數與 levels。同樣以 GuestNum 為例：



從合併 level 前的 Multiple Comparison 可以看到原本的 level 分類方式僅「單人」與「其他」有顯著差異，因此合併其他 levels，得到「單人」與「非單人」。(其他類別變數請詳 八、附錄)

#### 四、交互項

合併 Levels 後，本組認為單一變數不足以描述消費者樣態，為了提升模型解釋力，根據 domain knowhow 加入交互項如下：

Interaction Items	Interaction Items
<b>OrderPre:TravelDist</b>	消費者傾向提前多久訂房，可能會因旅行距離遠近而異。 例如：越是長距離旅行，會越早提前訂房。
<b>Season:TravelDist</b>	消費者偏好旅行距離之遠近，可能會因季節不同而異。 例如：偏好在夏季進行遠程旅行，在冬季進行國內旅行。
<b>Monetary:Area</b>	消費者每次旅行之平均花費，可能會因旅行地點而異。 例如：東北亞旅行之花費可能會高於東南亞旅行之花費。

由於變數之間可能有許多直覺上難以想到之交互關係，因此本組列出各種變數組合後，以 **step()** 函數篩選，最終得出 29 個交互項。(請詳 p.7 MLR Model 3)

#### 五、多元迴歸模型與廣義線性模型建置

##### 1. Multiple Linear Regression (MLR)

- ✓ 目的：預測顧客消費週期(T)
- ✓ 資料：篩選舊客(**Frequency > 1**)的客戶資料
- ✓ 模型建立過程：

###### ■ **Model 1**

**Y** : *TMed*

**X** : *TtlContri, Recency, Monetary, AvgPmt, DailyQpon, DailyNote, NightType, ChildType, GuestNum, Device, BnbType, DoW, OrderPre, Season, Area, TravelDist*

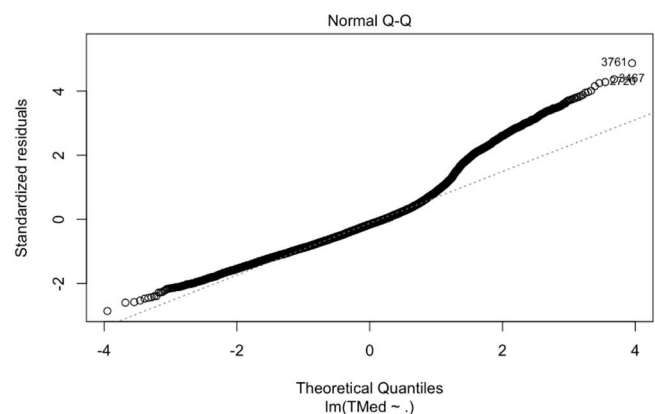
- (1) 首先對 **TMed** 做多元線性回歸結果如下：

Residual standard error: 111.6 on 12827 degrees of freedom  
Multiple R-squared: 0.3015, Adjusted R-squared: 0.2994  
F-statistic: 138.4 on 40 and 12827 DF, p-value: < 2.2e-16

R squared 為 0.3015，Residual standard error 為 111.6，模型的解釋力偏低。

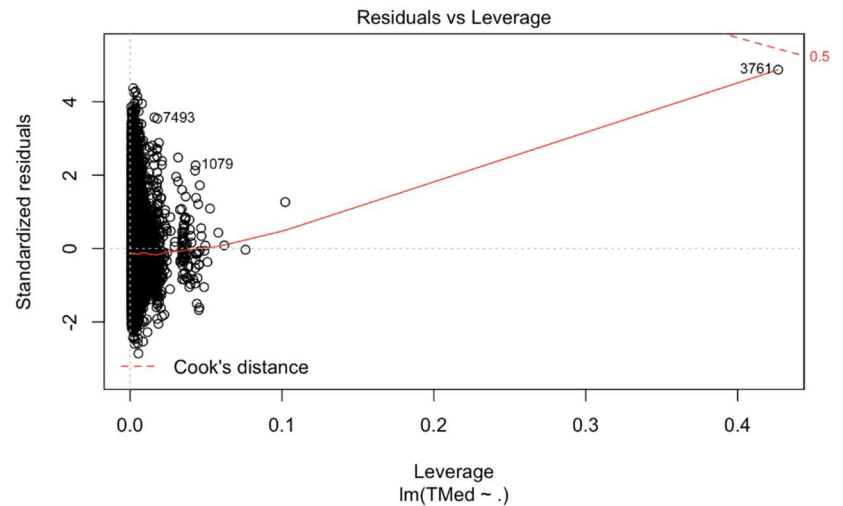
- (2) *Checking Model Assumptions* :

根據 **QQ-plot** 的結果，Model 1 違背了 Normality 的假設，Residuals 很明顯的偏離了虛線



## (3) Outliers

第 3761 筆資料的 **Cook's distance** 仍 **<0.5** 因此本組將保留該筆資料

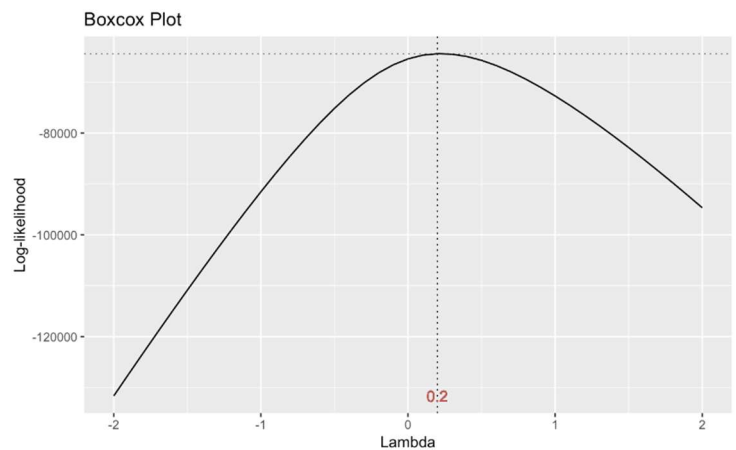
■ **Model 2**

$Y : \log(TMed)$

$X : TtlContri, Recency, Monetary, AvgPmt, DailyQpon, DailyNote, NightType, ChildType, GuestNum, Device, BnbType, DoW, OrderPre, Season, Area, TravelDist$

(1) 對 **TMed** 做 Box-Cox Transformation

根據 Box-Cox 的結果，likelihood 最大的 Lambda 的值約為 0.22。考量商業上解釋力，因 lambda 取 0.22 與取 lambda = 0 的結果差異不大，因此本組決定取 lambda=0，對 TMed 做 Log Transformation。

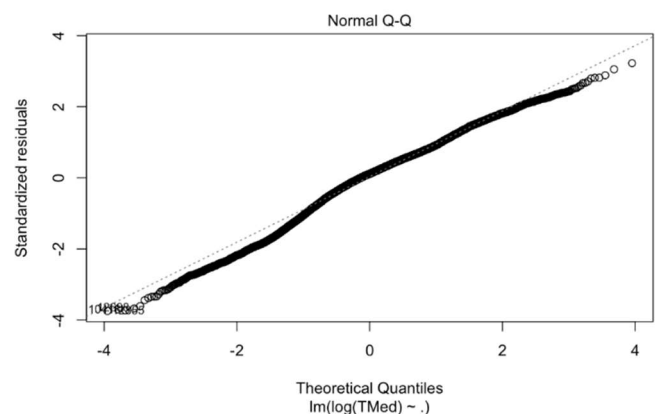


## (2) R-squared = 0.3777、Residual standard error = 1.426，模型解釋力仍偏低。

Residual standard error: 1.426 on 12826 degrees of freedom  
Multiple R-squared: 0.3777, Adjusted R-squared: 0.3757  
F-statistic: 194.6 on 40 and 12826 DF, p-value: < 2.2e-16

## (3) Checking Model Assumptions

經過 Log Transformation，**Model2** 符合線性迴歸的模型假設。



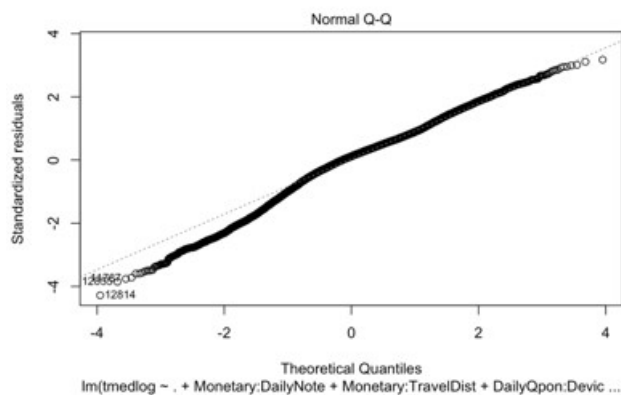
### ■ Model 3 (最終採用模型)

$Y : \log(TMed)$

$X : \text{原變數} + \text{交互項} :$

1. Monetary:DailyNote	8. GuestNum:OrderPre	15. Monetary:Area	22. DailyNote:GuestNum
2. Monetary:TravelDist	9. Device:DoW	16. Monetary:Season	23. DailyNote:Device
3. DailyQpon:Device	10. OrderPre:Season	17. AvgPmt:DailyQpon	24. DailyNote:Season
4. DailyQpon:BnbType	11. OrderPre:TravelDist	18. AvgPmt:Season	25. DailyNote:Area
5. DailyQpon:Season	12. Season:TravelDist	19. AvgPmt:Area	26. GuestNum:TravelDist
6. NightType:TravelDist	13. Monetary:AvgPmt	20. DailyQpon:DailyNote	27. Device:OrderPre
7. ChildType:Season	14. Monetary:GuestNum	21. DailyQpon:TravelDist	28. DoW:Season
			29. Area:TravelDist

#### (1) 符合模型假設



#### (2) 模型表現

	Model3	Model2
Residual standard error	1.383	1.432
R-squared	0.4203	0.3721
Adjusted R-squared	0.4133	0.3707

#### (3) 結論：採用 **Model 3** 預測顧客消費週期(T)

#### (4) 類別變數預期與實際影響比較表：

在實際進入迴歸模型前，本組預期各類別變數的 levels 會對 **TMed** 有不同程度的影響，因此以每個變數中沒有特別傾向 Level 為 Reference，以相對的概念呈現(透過 beta coefficient 判別)，用「+」表 **TMed** 相對 reference 增加、「-」表 **TMed** 相對 reference 減少、以及「△」表無法預期影響結果，並以符號數量表示影響程度大小，比對實際跑迴歸的影響，整理成下表，進行說明：

Variables	levels	預期對 TMed 影響	實際對 TMed 影響
NightType	most_1~3	-	-
	most_4up	+	+
	Other	Reference	

<b>ChildType</b>	<i>a_child</i>	+++	---
	<i>xa_child</i>	Reference	
<b>GuestNum</b>	<i>single</i>	-	-
	<i>nonsingle</i>	Reference	
<b>Device</b>	<i>app</i>	+	---
	<i>mobile</i>	+	--
	<i>pc</i>	++	-
	<i>neutral</i>	Reference	
<b>BnbType</b>	<i>apt</i>	-	+++
	<i>else.apt</i>	△	++
	<i>neutral</i>	Reference	
<b>DoW</b>	<i>weekday</i>	+	+
	<i>weekend</i>	+	+++
	<i>xspecial</i>	Reference	
<b>Season</b>	<i>spring</i>	+++	---
	<i>summer</i>	+++	-
	<i>autumn</i>	+++	-
	<i>winter</i>	+++	+++
	<i>xspecial</i>	Reference	
<b>OrderPre</b>	<i>near</i>	Reference	
	<i>normal</i>	++	---
	<i>normal+</i>	++	+++
	<i>early</i>	+++	---
	<i>early+</i>	+++	-
<b>Area</b>	<i>EastAsia</i>	△	++
	<i>EastSouthAsia</i>	△	+
	<i>Others</i>	Reference	
<b>Travel_dist</b>	<i>P_Local</i>	---	---
	<i>P_Short</i>	-	+++
	<i>P_Far</i>	+++	+++
	<i>No_Tendency</i>	Reference	

- **NightType**：推測住的時間愈短的為小旅遊或出差，頻率會較高，因此相較沒特別偏好的 T 低
- **ChildType**：有小孩的為家庭旅遊，通常頻率會較低，因此 T 高
- **GuestNum**：推測單人旅遊比較好規劃且時間彈性，也可能為出差，因此頻率會比較高，T 低
- **Device**：推測用多管道會讓訂房更為方便並能更頻繁訂購，加上電腦螢幕大，比較方便看，因此其他裝置對 T 的影響相對為負
- **BnbType**：推測住 apt 的為小資旅遊，旅遊頻繁能較為頻繁，T 相較沒特別偏好的顧客應該較低
- **DoW**：推測沒特別偏好時間的顧客會比較常出去玩，較多隨興小旅遊，因此 T 低；會特地選時間訂房的是較有規劃的顧客，為較長期的旅遊，頻率會比較低，因此 T 高
- **Season**：會偏好某季節出遊的顧客，下次旅行會等一年到偏好的季節，因此 T 高
- **OrderPre**：推測時間很接近才訂房的為短程旅遊或國內旅遊，沒有急迫性，加上短程旅遊的頻率會比長程旅遊高，因此 T 高
- **Area**：無法預料結果會如何
- **Travel\_dist**：推測旅遊距離愈遠，頻率與低，T 愈高，因為旅遊時間會較長且金額較高



針對預期與實際差異較大的變數與 levels，以下分析可能原因：

- **ChildType**：原先預期帶小孩出遊的家庭旅遊頻率會較低，然而結果相反，分析可能原因為現在許多家庭會趁小孩還小多出遊以創造回憶以及珍惜與家人同遊的時間，因此實際上 T 較沒有帶小孩出遊低。
- **Device**：原先預期使用多管道的用戶 T 會較高，然而結果相反，分析原因為現今年輕人多習慣用手機，因此多管道並沒有帶來更多便利性，也因此實際上 T 較高。
- **Season**：原先預期偏好特定季節出遊的用戶有較長的周期，實際結果看不出特別偏好。
- **OrderPre**：原先預期愈接近出遊時間訂房的為短期旅遊，較頻繁，然而實際結果看不出有特定規律。

## 2. Generalized linear model (GLM)

- ✓ 目的：預測顧客存活率
- ✓ 資料：篩選舊客(*Frequency > 1*)的客戶資料
- ✓ 模型架構與定義：



### ■ Asiayo 顧客屬於「非契約顧客」

### ■ Status 定義：

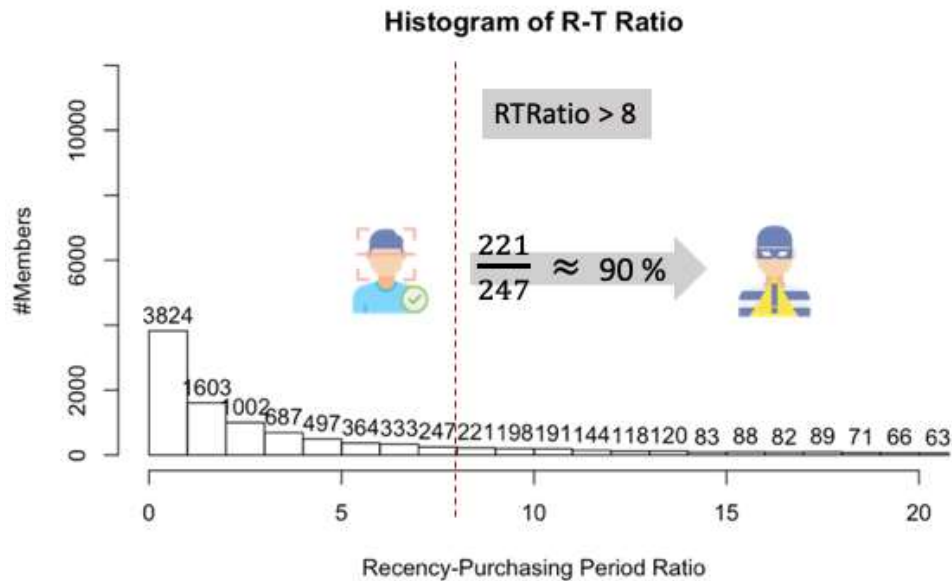
Status 是 GLM 模型所要看的被解釋變數，本組希望能夠去看每一位顧客對於 Asiayo 平台的存活與否，因此定義當 Status = 1 時為存活，反之當 Status = 0 時代表顧客將離開平台(死亡)。而衍生出 Status 這個新變數的方法則是根據 RTRatio 的比率計算。

### ■ RTRatio

$RTRatio = Recency / TMed$ ，由於本組假設每位顧客的回購行為都存在自己的購物週期，因此採用  $Recency / T$  的比值作為評斷標準，當顧客  $Recency / T$  比值越接近 1 時，代表該顧客回購機率越高；而當  $Recency / T$  比值太大，代表該顧客可能已經想離開或已經離開 AsiaYo。







上圖為 RTRatio 的分佈圖，本組去計算相鄰 RTRatio 的比值(ex. RTRatio 2/ RTRatio 1)，當比值 > 90% 本組判定該顧客未來只會跳到更高的 RTRatio (往圖的右方跳)，而其跳回較低的 RTRatio 的機率則很低。本組計算出當 RTRatio = 8 時，顧客往後移轉的機率為 90%，因此定義當某顧客的 RTRatio > 8 時判定該顧客死亡，反之存活。根據此判斷邏輯定義出衍生變數 Status，並對其建構 GLM 模型。

#### ■ 模型結果

**Y : Status**

**X : Monetary、DailyQpon、DailyNote、ChildType、GuestNum、Device、BnbType、DoW、OrderPre、Season、Area、TravelDist、多項 Interaction Terms**

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 16411 on 12867 degrees of freedom  
Residual deviance: 12678 on 12800 degrees of freedom  
AIC: 12814

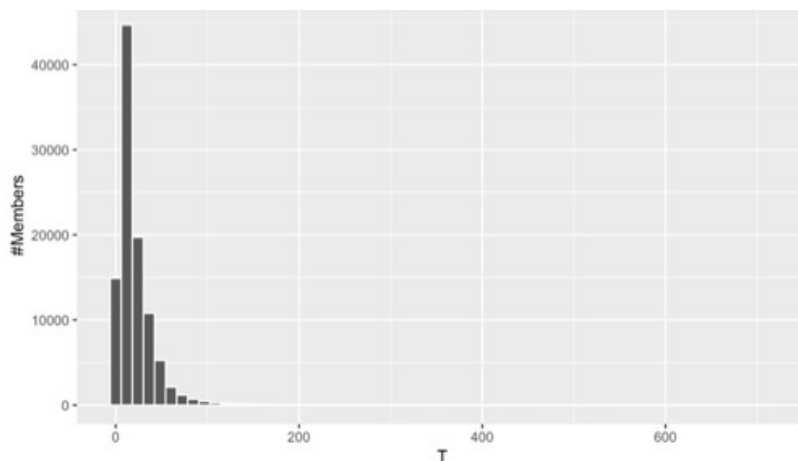
```
```{r}
(1-pchisq(deviance(logit0), df.residual(logit0)))
```
```

[1] 0.7760025

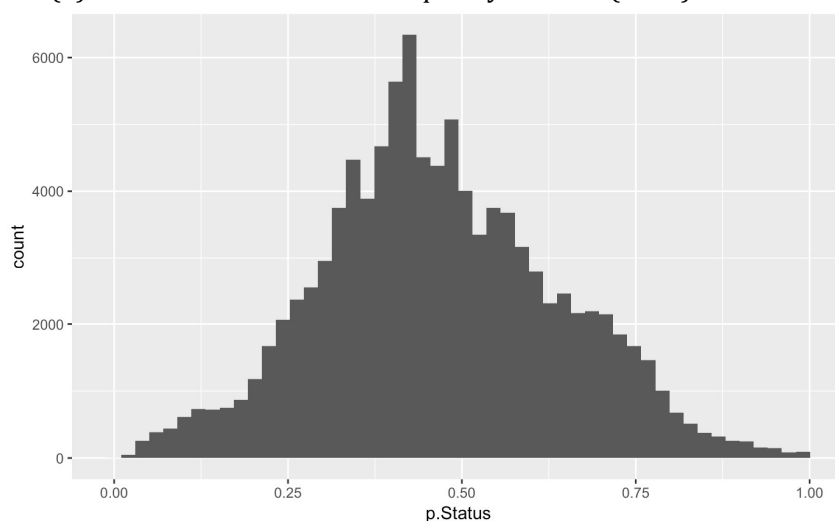
- (1) Deviance 從 Null Deviance 的 16411 降到 12678，下降的幅度為 3544 代表模型所加入的解釋變數是具有解釋能力的
- (2) Deviance 與 DF 之間的差距不大
- (3) 對模型做 Good Fit Model Test 中的 Likelihood Ratio Test，得到 p-value 為 0.776，接受虛無假設，因此判定模型是好的

|                                     | Estimate   | Std. Error | z value | Pr(> z ) |    |                                  |            |           |
|-------------------------------------|------------|------------|---------|----------|----|----------------------------------|------------|-----------|
| (Intercept)                         | -2.028e+00 | 7.050e-01  | -2.876  | 0.004024 | ** |                                  |            |           |
| Monetary                            | 7.779e-05  | 2.484e-05  | 3.132   | 0.001735 | ** |                                  |            |           |
| DailyQpon                           | 5.943e-01  | 1.479e-01  | 4.017   | 5.89e-05 | ** |                                  |            |           |
| DailyNote                           | -1.002e-01 | 6.104e-02  | -1.642  | 0.100543 |    |                                  |            |           |
| ChildTypexa_child                   | 8.868e-02  | 3.395e-01  | 0.261   | 0.793950 |    | DoWeekend:Seasonno_special       | -2.405e-01 | 2.161e-01 |
| GuestNumsingl                       | -4.284e-01 | 8.467e-02  | -5.059  | 4.22e-07 | ** | DoWxspecial:Seasonno_special     | 1.090e-01  | 1.703e-01 |
| Devicemobile                        | -2.332e-01 | 6.466e-02  | -3.607  | 0.000310 | ** | DoWeekend:Seasonspring           | 3.086e-02  | 2.291e-01 |
| Deviceneutral                       | 8.738e-02  | 6.814e-02  | 1.282   | 0.199714 |    | DoWxspecial:Seasonspring         | -4.300e-01 | 1.931e-01 |
| Devicepc                            | -2.239e-01 | 6.245e-02  | -3.586  | 0.000336 | ** | DoWeekend:Seasonsummer           | -1.392e-02 | 2.070e-01 |
| BnbTypeeelse.ap                     | -9.738e-02 | 7.458e-02  | -1.306  | 0.191643 |    | DoWxspecial:Seasonsummer         | -3.207e-01 | 1.704e-01 |
| BnbTyenutral                        | -1.143e-01 | 5.846e-02  | -1.956  | 0.050480 |    | DoWeekend:Seasonwinter           | -2.689e-01 | 2.332e-01 |
| DoWeekend                           | 2.718e-01  | 1.804e-01  | 1.506   | 0.131972 |    | DoWxspecial:Seasonwinter         | -5.541e-01 | 2.000e-01 |
| DoWxspecial                         | -1.243e-01 | 1.435e-01  | -0.866  | 0.386374 |    | Monetary:Seasonno_special        | -4.473e-05 | 2.886e-05 |
| OrderPrearily+                      | 9.920e-01  | 6.557e-01  | 1.513   | 0.130309 |    | Monetary:Seasonspring            | 1.530e-05  | 3.019e-05 |
| OrderPrear                          | 2.651e+00  | 5.871e-01  | 4.516   | 6.31e-06 | ** | Monetary:Seasonsummer            | -1.750e-06 | 2.785e-05 |
| OrderPrenormal                      | 7.057e-01  | 6.212e-01  | 1.136   | 0.255977 |    | Monetary:Seasonwinter            | -7.280e-06 | 3.116e-05 |
| OrderPrenormal+                     | 1.058e+00  | 6.301e-01  | 1.680   | 0.093024 |    | DailyQpon:Seasonno_special       | 5.175e-02  | 1.943e-01 |
| Seasonno_special                    | 3.658e+00  | 9.524e-01  | 3.841   | 0.000123 | ** | DailyQpon:Seasonspring           | 4.756e-01  | 2.202e-01 |
| Seasonspring                        | -1.292e+00 | 9.727e-01  | -1.329  | 0.183957 |    | DailyQpon:Seasonsummer           | 2.948e-01  | 1.858e-01 |
| Seasonsummer                        | -4.319e-01 | 8.014e-01  | -0.539  | 0.589951 |    | DailyQpon:Seasonwinter           | 4.648e-01  | 2.280e-01 |
| Seasonwinter                        | 8.653e-01  | 8.694e-01  | 0.995   | 0.319552 |    | OrderPrearily+:Seasonno_special  | -1.305e+00 | 8.947e-01 |
| AreaEastSouthAsia                   | -1.373e+00 | 1.614e-01  | -8.505  | < 2e-16  | ** | OrderPrear+:Seasonno_special     | -2.509e+00 | 7.921e-01 |
| AreaOthers                          | -8.537e-01 | 2.814e-01  | -3.034  | 0.002410 | ** | OrderPrenormal+:Seasonno_special | -7.237e-01 | 8.270e-01 |
| TravelDist2P_Local                  | -2.414e-01 | 4.318e-01  | -0.559  | 0.576165 |    | OrderPrenormal+:Seasonspring     | -1.192e+00 | 8.454e-01 |
| Seasonno_special:TravelDist2P_Local | 1.815e-01  | 2.622e-01  | 0.692   | 0.488918 |    | OrderPrearily+:Seasonspring      | 1.119e-01  | 7.914e-01 |
| Seasonspring:TravelDist2P_Local     | 2.018e-01  | 2.764e-01  | 0.730   | 0.465281 |    | OrderPrear+:Seasonspring         | -1.406e+00 | 6.827e-01 |
| Seasonsummer:TravelDist2P_Local     | 3.339e-01  | 2.548e-01  | 1.310   | 0.190131 |    | OrderPrenormal+:Seasonspring     | -5.817e-01 | 7.142e-01 |
| Seasonwinter:TravelDist2P_Local     | -2.152e-02 | 2.803e-01  | -0.077  | 0.938821 |    | OrderPrenormal+:Seasonsummer     | -9.822e-01 | 7.351e-01 |
| ChildTypexa_child:Seasonno_special  | 1.004e-01  | 4.835e-01  | 0.208   | 0.835478 |    | OrderPrearily+:Seasonsummer      | -7.464e-01 | 7.410e-01 |
| ChildTypexa_child:Seasonspring      | 1.448e+00  | 6.372e-01  | 2.273   | 0.023040 | *  | OrderPrear+:Seasonsummer         | -1.311e+00 | 6.529e-01 |
| ChildTypexa_child:Seasonsummer      | 1.001e+00  | 4.011e-01  | 2.496   | 0.012552 | *  | OrderPrenormal+:Seasonsummer     | -7.156e-01 | 6.776e-01 |
| ChildTypexa_child:Seasonwinter      | 6.081e-01  | 4.786e-01  | 1.271   | 0.203846 |    | OrderPrenormal+:Seasonwinter     | -1.312e+00 | 6.961e-01 |
| OrderPrearily+:TravelDist2P_Local   | 1.622e-01  | 5.153e-01  | 0.315   | 0.752966 |    | OrderPrear+:Seasonwinter         | -1.096e+00 | 7.647e-01 |
| OrderPrear+:TravelDist2P_Local      | 2.527e-02  | 3.776e-01  | 0.067   | 0.946636 |    | OrderPrenormal+:Seasonwinter     | -2.134e+00 | 6.695e-01 |
| OrderPrenormal+:TravelDist2P_Local  | 1.105e+00  | 4.171e-01  | 2.650   | 0.008041 | ** | OrderPrear+:Seasonwinter         | -1.407e+00 | 7.014e-01 |
| OrderPrenormal+:TravelDist2P_Local  | 6.081e-01  | 4.149e-01  | 1.466   | 0.142735 |    | OrderPrenormal+:Seasonwinter     | -1.470e+00 | 7.168e-01 |

(4) 應用 MLR model 預測 Frequency = 1 的消費週期(T)



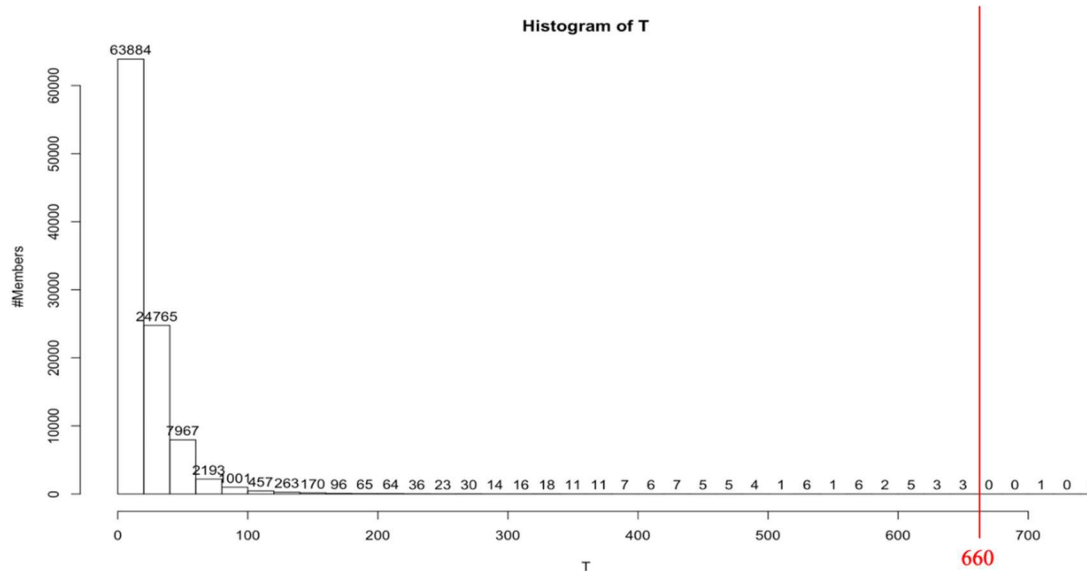
(5) 應用 GLM model 預測 Frequency = 1 的 P(Alive)機率分佈

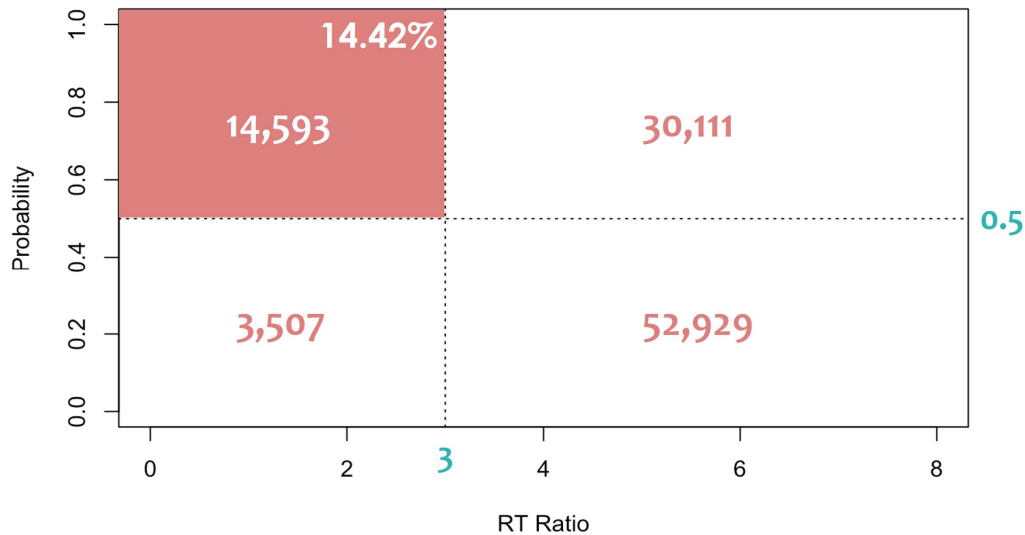


## 六、預測結果與行銷建議

### 1. 目標客群選擇

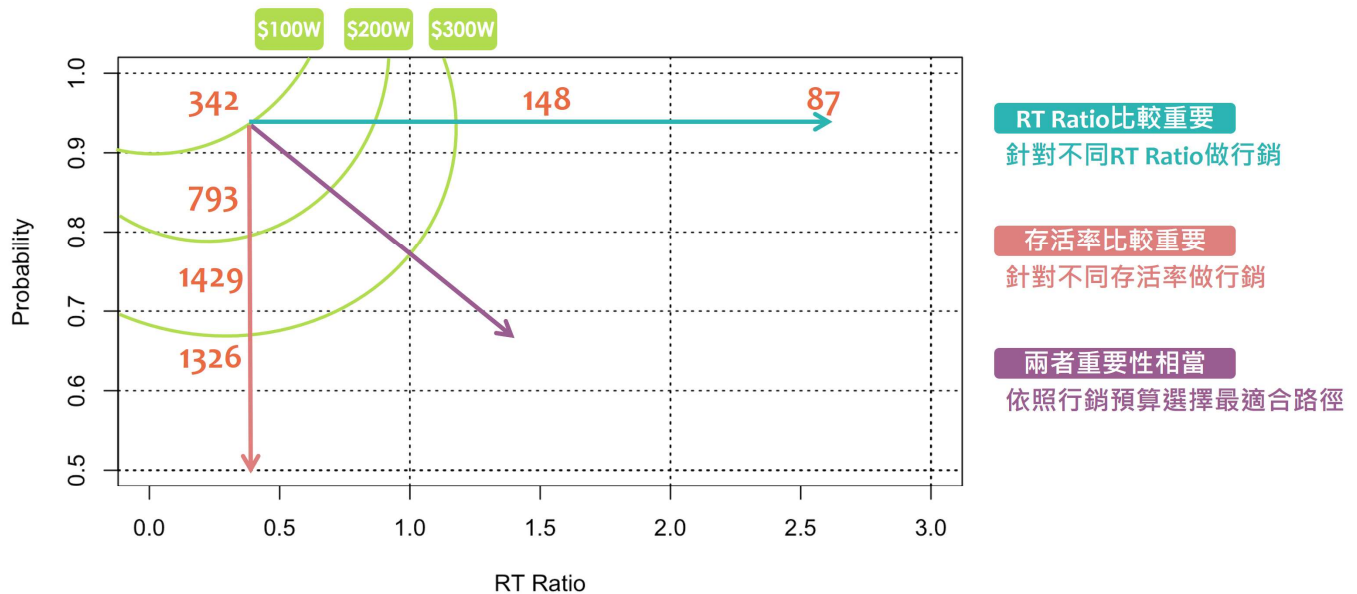
為了能夠將資源做最妥善的利用，本組認為必須將目標客群做一個清楚的定義。其中，本組的客群首先設定客群為新客 (**Frequency=1**)，並且利用前面提到的預測方式預測 TMed 數值。然而，AsiaYo 提供的資料中 TMed 最大值為 660 天，為了避免外插所造成的可能偏誤，本組將回購週期(T)大於 660 天者，視為該顧客流失。其人數的分配如下圖所示：





## 2. 行銷方案實行

在決定了目標客群之後，本組提供了 AsiaYo 一些未來可行的方向，AsiaYo 可以從 RT Ratio 亦可以從存活率下手，取決於 AsiaYo 對於這兩者重要性的排序。首先，如果認為 RT Ratio 比較重要，那便可以針對不同 RT Ratio 做行銷（如下圖中的藍色箭頭左至右）；如果認為存活率比較重要，則可以針對不同存活率做行銷（如下圖中的紅色箭頭上至下）；再者，如果認為兩者重要性相當，則依照下圖中行銷預算選擇最適合路徑（依照下圖的紫色箭頭配合假定為不同行銷預算的紫色扇形從左上往右下）。圖如下：



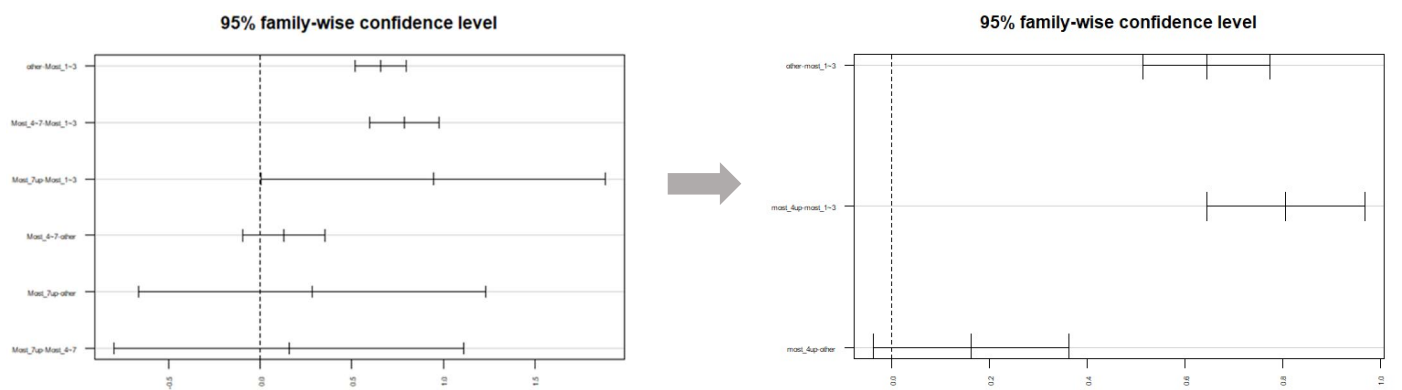
## 七、總結

本組總結報告要點如下：

1. 使用 RT Ratio 與存活率取代某固定時間的 Retention rate
2. 使用 One Number One Row 的型態 ( 將 Order ID 轉換成 User ID )
3. 以 MLR 模型預測新客『回購週期 T』
4. 以 GLM 模型預測新客『存活率』
5. 依照行銷預算選擇適當的路徑
  - (1) RT Ratio 比較重要：行銷方案投遞從 RT Ratio 低到高
  - (2) 存活率比較重要：行銷方案投遞從存活率高到低
  - (3) 兩者並重：依照行銷預算選擇最適合路徑

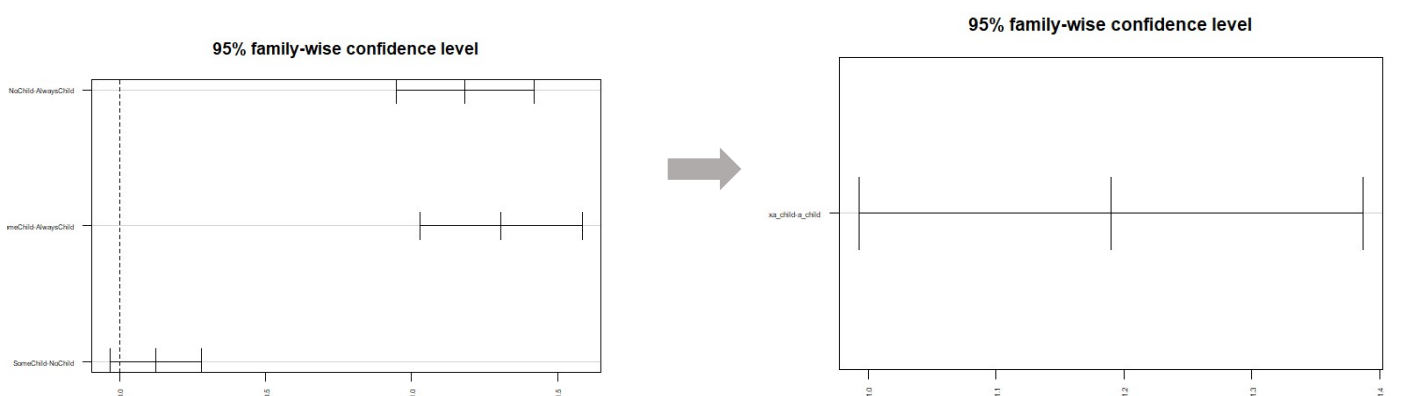
## 八、附錄

### 1. NightType 篩選邏輯：



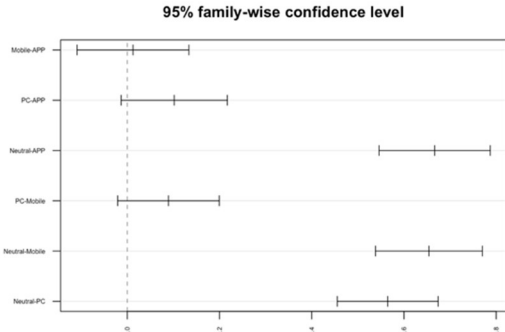
| Levels of "NightType" | Rules                   |
|-----------------------|-------------------------|
| <i>most_1~3</i>       | 訂房一至三晚的比例大於整體訂房 50%     |
| <i>most_4up</i>       | 訂房四晚以上(包含四)比例大於整體訂房 50% |
| <i>other</i>          | 其他不包含以上分類者              |

### 2. ChildType 篩選邏輯：



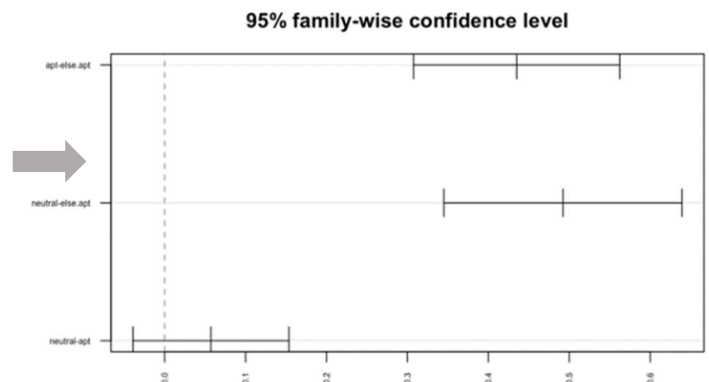
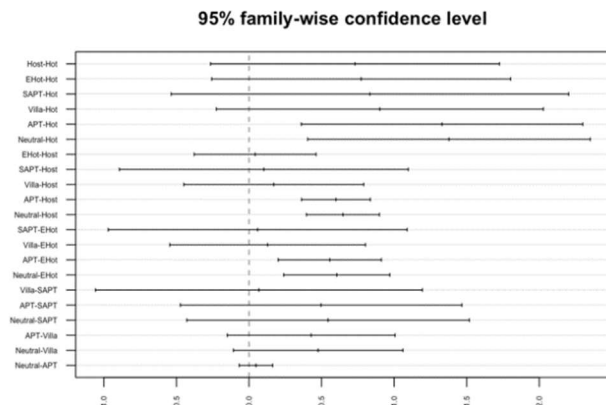
| <i>Levels of "ChildType"</i> | <i>Rules</i>  |
|------------------------------|---------------|
| <i>a_child</i>               | User 每次訂房都有小孩 |
| <i>xa_child</i>              | 其他不包含以上者      |

### 3. Device 篩選邏輯 (照舊) :



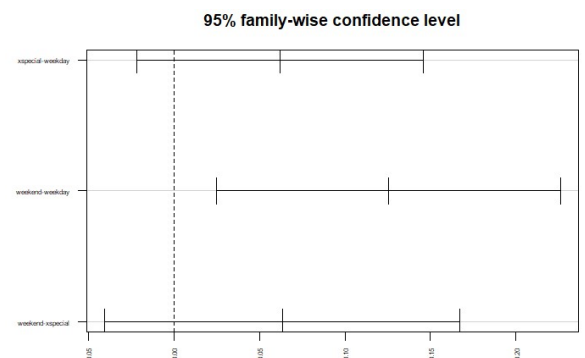
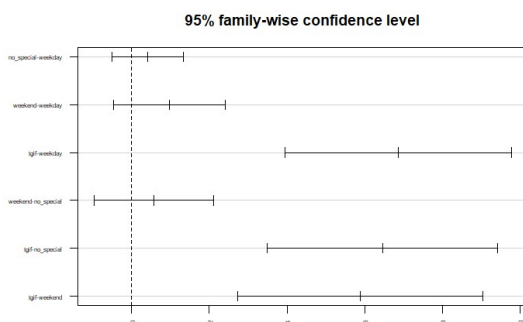
| <i>Levels of "Device"</i> | <i>Rules</i>             |
|---------------------------|--------------------------|
| <i>app</i>                | 使用 iOS+Android 的比例大於 50% |
| <i>mobile</i>             | 使用 Mobile 的比例大於 50%      |
| <i>pc</i>                 | 使用 PC 的比例大於 50%          |
| <i>neutral</i>            | 其他不包含以上分類者               |

### 4. BnbType 篩選邏輯 :



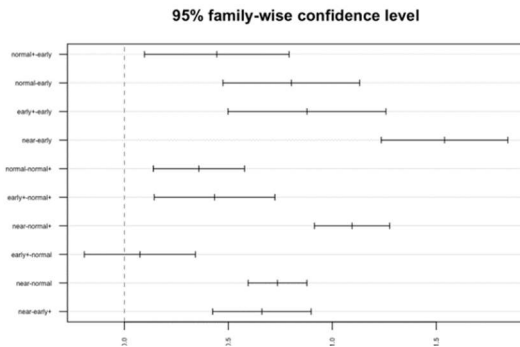
| <i>Levels of "BnbType"</i> | <i>Rules</i>   |
|----------------------------|--|
| <i>apt</i>                 | 訂購房型為 Apartment 比例大於 50%                               |
| <i>else.apt</i>            | 訂購房型為 Economy Hotel + Hostel+Hotel+SAPT+Villa 比例大於 50% |
| <i>neutral</i>             | 其他不包含以上分類者   |

### 5. DoW (the Day of the Week)篩選邏輯 :



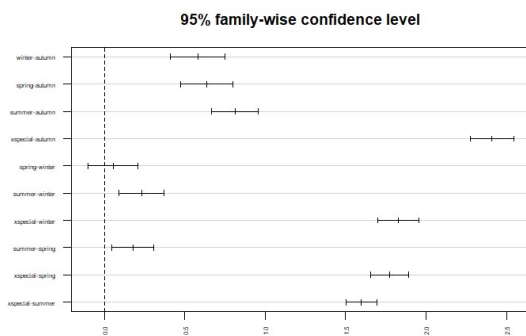
| <b>Levels of "DoW"</b> | <b>Rules</b>             |
|------------------------|--------------------------|
| <i>weekday</i>         | 平日訂房佔顧客整體比例大於 50%        |
| <i>weekend</i>         | 周末(包含星期五)訂房佔顧客整體比例大於 50% |
| <i>xspecial</i>        | 其他不包含以上分類者               |

## 6. OrderPre 篩選邏輯 ( 照舊 ):



| <b>Levels of "OrderPre"</b> | <b>Rules</b>              |
|-----------------------------|---------------------------|
| <i>supernear</i>            | 在入住的 3 天以內訂房比例大於 50%      |
| <i>near</i>                 | 在入住前的 4-13 天期間訂房比例大於 50%  |
| <i>normal</i>               | 在入住前的 14-30 天期間訂房比例大於 50% |
| <i>normal+</i>              | 在入住前的 31-60 天期間訂房比例大於 50% |
| <i>early</i>                | 在入住前的 61-90 天期間訂房比例大於 50% |
| <i>early+</i>               | 在入住前的 91 天以上期間訂房比例大於 50%  |
| <i>neural</i>               | 其他不包含以上分類者                |

## 7. Season 篩選邏輯 ( 照舊 ):

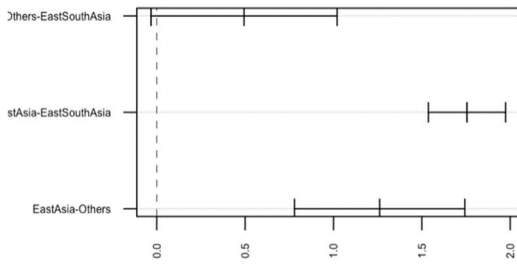


| <b>Levels of "Season"</b> | <b>Rules</b>  |
|---------------------------|---|
| <i>spring</i>             | 於三四五月份入住比例大於 50%                                    |
| <i>summer</i>             | 於六七八九月份入住比例大於 50%<br>時間拉長為四個月原因：國人暑假出遊時間通常是從六月底到九月初 |
| <i>autumn</i>             | 於十十一十二月份入住比例大於 50%                                  |
| <i>winter</i>             | 於十二一二月份入住比例大於 50%<br>此部分也可視為國人春節與寒假出遊習慣             |
| <i>xspecial</i>           | 其他不包含以上分類者  |



## 8. Area 篩選邏輯 ( 照舊 ):

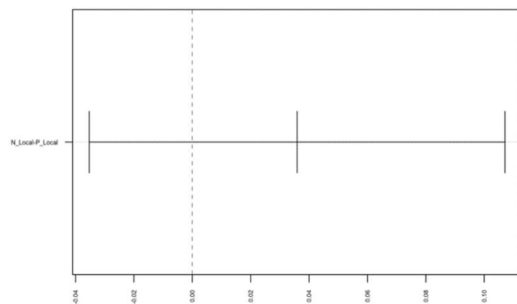
95% family-wise confidence level



| Levels of "Area" | Rules   |
|------------------|---|
| EastAsia         | Taiwan, Japan, Korea, Hong Kong, Macau, Korea-North |
| EastSouthAsia    | Malaysia, Singapore, Thailand, Vietnam              |
| Others           | Other country (Europe, America)                     |

## 9. Travel\_dist 篩選邏輯 ( 照舊 ):

95% family-wise confidence level



| Levels of "Area" | Rules   |
|------------------|---|
| P_Local          | 出發地和目的地相同的比例大於 50%  |
| P_Short          | 出發地和目的地為短程( EastAsia 到 EastAsia 或 EastSouthAsia 到 EastSouthAsia ) 的比例大於 50% |
| P_Medium         | 出發地和目的地為中程( EastAsia 到 EastSouthAsia 或 EastSouthAsia 到 EastAsia ) 的比例大於 50% |
| P_Long           | 出發地和目的地為長程( Others 到 EastSouthAsia 或 EastAsia ) 的比例大於 50%                   |
| No Tendency      | 其他不包含以上分類者  |
| N_Local          | 不包含在 P_Local 者  |