題目：

阿竹是一個旅館銷售人員，負責某間旅館的銷售業務，他想針對旅館網頁的訪客進行廣告優惠，但是行銷經費有限，只能針對較有可能消費的訪客進行廣告。阿竹提供了旅館網站的Google Analytics流量統計資料，請資料工程師幫助阿竹找出可能進行消費的訪客。

1. 基本題：預測訪客是否會訂房，請依照流量統計資料train.csv，訓練一個分類模型或回歸模型，預測test.csv中每位訪客是否會進行消費的機率，並依照sample\_submission.csv的格式提交預測結果。請記錄資料分析、清理、建模、模型評估的過程，將完整過程、清理後資料與相關程式碼上傳至個人github並提供連結或是寄至 adrian.chen@gmail.com

\* 數據包含10個數字屬性和8個分類屬性

\* 'Revenue' 屬性將作為標籤 1 代表會訂房 0 代表不會

\* 使用訓練資料 train.csv 建立模型，預測 test.csv 是否會訂房

\* 屬性說明：

- Administrative：管理頁面瀏覽數量

- Administrative\_Duration：管理頁面瀏覽時間

- Informational：資訊頁面瀏覽數量

- Informational\_Duration：資訊頁面瀏覽時間

- ProductRelated：產品頁面瀏覽數量

- ProductRelated\_Duration：產品頁面瀏覽時間

- BounceRates：跳出率

- ExitRates：退出率

- PageValues：頁面價值

- SpecialDay：特殊日期（例如聖誕節，情人節）接近程度

- Month：訪問月份

- OperatingSystems：作業系統

- Browser：瀏覽器

- Region：地區  
 - TrafficType：流量來源

- VisitorType：訪客類型

- Weekend：訪問時間是否為週末

- Revenue：是否訂房

2. 加分題：延續上述案例，請依照以下二個混亂矩陣，評估何者較適合用於本案例的情境，將分析過程上傳至個人的github並提供連結或是寄至 adrian.chen@gmail.com

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model A | | | Model B | | |
|  | 預測為真 | 預測為假 |  | 預測為真 | 預測為假 |
| 實際為真 | 853 | 576 | 實際為真 | 846 | 583 |
| 實際為假 | 341 | 7230 | 實際為假 | 316 | 7255 |

模型評估 混亂矩陣比較