

# 智慧型飯店 行銷管理預測系統

指導老師: 蔡智勇老師 . 黃登揚老師

2020/05/05



勞動部勞動力發展署桃竹苗分署

109年度「亞洲·矽谷」產業專業人才發展基地

建構創新訓練模式 培育科技專業人才



財團法人自強工業科學基金會



# 組員介紹



- 組員: 許淑玲
- 專長:
- 負責: 網頁製作



- 組員: 李宜鏞 Greg
- 專長: 光通訊模組研發
- 負責: 推薦房型預測分項



- 組長: 李紹豪 James
- 專長: 車用電子產品PM
- 負責: 客戶忠誠度預測分項



- 組員: 張齡元 Lily
- 專長:
- 負責: 預定取消預測分項



- 組員: 張維倫
- 專長:
- 負責: 來客數預測分項



- 組員: 林承融
- 專長: 化學系
- 負責: Hadoop 分散式平台架設, 資料清洗



# TEAM

# 專題報告大綱



1. 資料來源與描述
2. 專題分項介紹
3. 應用與結論





# 1.資料來源與描述



# 資料背景

- 從 Kaggle 網站上尋找適合的題目及資料後下載進行專題研究分析。
- 此資料是兩個帶有酒店需求數據的數據集，其中一家酒店（H1）是度假酒店，另一家酒店是城市酒店（H2）。
- 兩個數據集共享相同的結構，其中31個變量描述了（H1）的40,060個觀測值和（H2）的79,330個觀測值，每個觀察值代表一個酒店預訂。
- 這兩個數據集均包含於2015年7月1日至2017年8月31日之間到達的預訂，包括有效完成的預訂和已取消的預訂。由於這是酒店的真實數據，因此刪除了與酒店或客戶標識有關的所有數據元素。

# H1 dataset summary statistics



| Variable                       | Min        | Max        | Median     | Unique |
|--------------------------------|------------|------------|------------|--------|
| <b>Reservation Status Date</b> | 2014-11-18 | 2017-09-14 | 2016-07-31 | 913    |

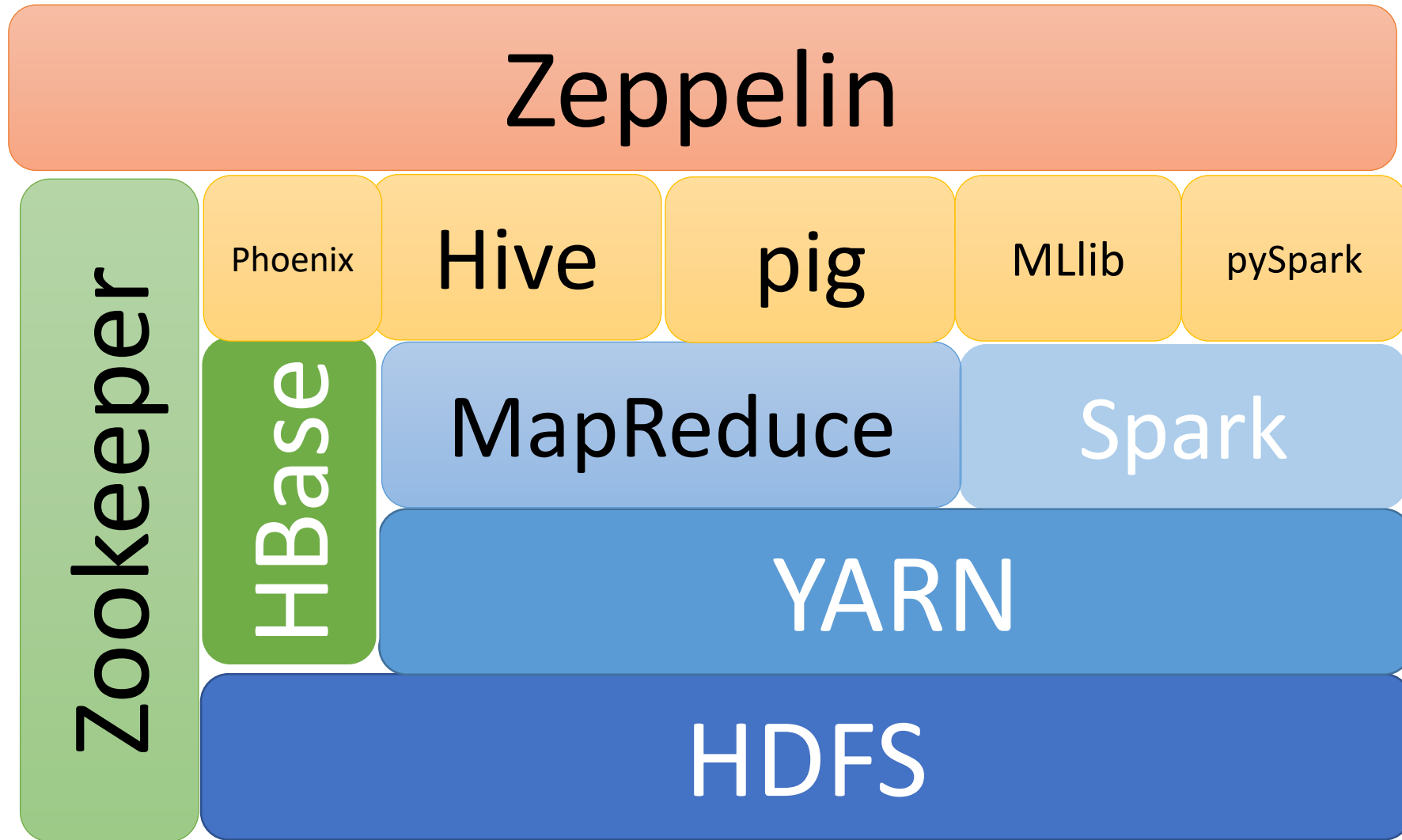
Table 2. Date variables

| Variable                    | Unique | Top counts   |
|-----------------------------|--------|--|
| <b>Agent</b>                | 186    | 240: 13 095, NULL: 8 209, 250: 2 869, 241: 1 721       |
| <b>Arrival Date Month</b>   | 12     | Aug: 4 894, Jul: 4 573, Apr: 3 609, May: 3 559         |
| <b>Assigned Room Type</b>   | 11     | A: 17 046, D: 10 339, E: 5 638, C: 2 214               |
| <b>Company</b>              | 236    | NULL: 36 952, 223: 784, 281: 138, 154: 133             |
| <b>Country</b>              | 125    | PRT: 17 630, GBR: 6 814, ESP: 3 957, IRL: 2 166        |
| <b>Customer Type</b>        | 4      | Tra.: 30 209, Tra.-Party: 7 791, Con.: 1 776, Gro.:284 |
| <b>Deposit Type</b>         | 3      | No Dep.: 38 199, Non-Refund.: 1 719, Ref.: 142         |
| <b>Distribution Channel</b> | 4      | TA/TO: 28 295, Dir.: 7 865, Cor.: 3 269, Und.: 1       |
| <b>Is Canceled</b>          | 2      | 0: 28 938, 1: 11 122                                   |
| <b>Is Repeated Guest</b>    | 2      | 0: 38 282, 1: 1 778                                    |
| <b>Market Segment</b>       | 6      | Onl.: 17 729, Off.: 7472, Dir.: 6 513, Gro.: 5 836     |
| <b>Meal</b>                 | 5      | BB: 30 005, HB: 8 046, Und.: 1 169, FB: 754            |
| <b>Reservation Status</b>   | 3      | C.Out: 28 938, Can.: 10 831, No-Show: 291              |
| <b>Reserved Room Type</b>   | 10     | A: 23 399, D: 7 433, E: 4 892, G: 1610                 |

Table 3. Categorical variables

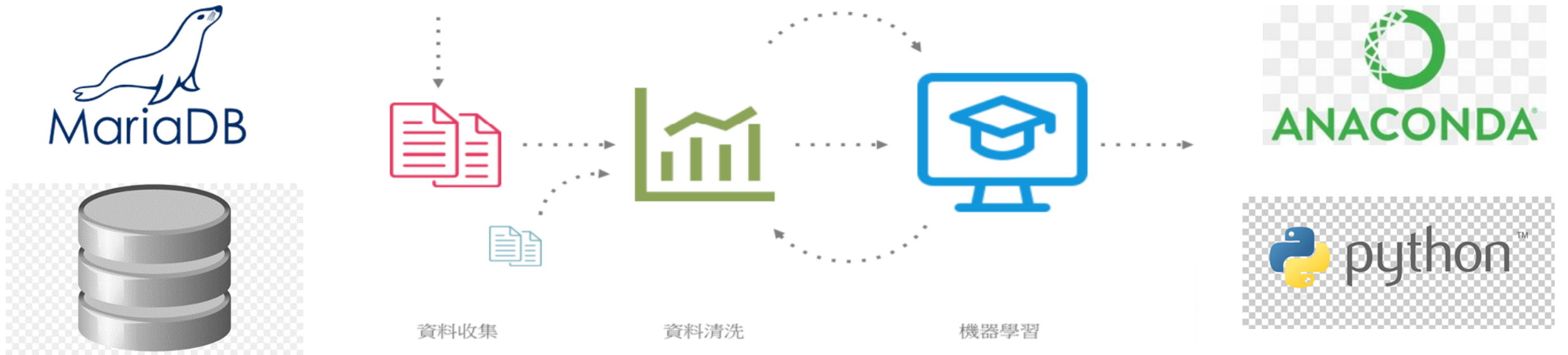
| Variable                              | Mean    | SD    | P0    | P25  | Median | P75  | P100 |
|---------------------------------------|---------|-------|-------|------|--------|------|------|
| <b>ADR</b>                            | 94.95   | 61.44 | -6.38 | 50   | 75     | 125  | 508  |
| <b>Adults</b>                         | 1.87    | 0.7   | 0     | 2    | 2      | 2    | 55   |
| <b>Arrival Date Of Month</b>          | 15.82   | 8.88  | 1     | 8    | 16     | 24   | 31   |
| <b>Arrival Date Week Number</b>       | 27.14   | 14.01 | 1     | 16   | 28     | 38   | 53   |
| <b>Arrival Date Year</b>              | 2016.12 | 0.72  | 2015  | 2016 | 2016   | 2017 | 2017 |
| <b>Babies</b>                         | 0.014   | 0.12  | 0     | 0    | 0      | 0    | 2    |
| <b>Booking Changes</b>                | 0.29    | 0.73  | 0     | 0    | 0      | 0    | 17   |
| <b>Children</b>                       | 0.13    | 0.45  | 0     | 0    | 0      | 0    | 10   |
| <b>Days In Waiting List</b>           | 0.53    | 7.43  | 0     | 0    | 0      | 0    | 185  |
| <b>Lead Time</b>                      | 92.68   | 97.29 | 0     | 10   | 57     | 155  | 737  |
| <b>Previous Bookings Not Canceled</b> | 0.15    | 1     | 0     | 0    | 0      | 0    | 30   |
| <b>Previous Cancellations</b>         | 0.1     | 1.34  | 0     | 0    | 0      | 0    | 26   |
| <b>Required Car Parking Spaces</b>    | 0.14    | 0.35  | 0     | 0    | 0      | 0    | 8    |
| <b>Stays In Weekend Nights</b>        | 1.19    | 1.15  | 0     | 0    | 1      | 2    | 19   |
| <b>Stays In Week Nights</b>           | 3.13    | 2.46  | 0     | 1    | 3      | 5    | 50   |
| <b>Total Of Special Requests</b>      | 0.62    | 0.81  | 0     | 0    | 0      | 1    | 5    |

Table 4. Integer and numeric variables





# 數據分析流程暨開發語言軟體







# 問題與對策

- 根據資料，我們發現淡季、旺季的來客數是影響飯店營收的一個重要因素，因此如何讓飯店提高來客數成為我們研究的重點。
- 要提高來客數，首先須先預測客戶入住的日期以便提早安排餐點，房型，停車，設施，房型等等相關細節。
- 因此進行大數據分析後，找出各國節假日，歷年來客數，星期平日、假日，淡季、旺季等關鍵因子，提供飯店AI預測結果。
- 再者提高客戶忠誠度也可增加來客數。利用機器學習關聯分析技術找出關鍵因子，提供飯店訂房管理人員及時預測網路訂房客戶的忠誠度，作為提供獎勵折扣的依據及旺季客戶的訂房優先順序的判定。



# 問題與對策

- 訂房後退訂常常會造成飯店管理與安排上的混亂，因此造成了飯店為了提高住房率而出現超賣房間的情形。
- 依照客戶輸入的訂房條件，利用機器學習關聯分析技術找出關鍵因子，提供飯店訂房管理人員及時預測網路訂房客戶的退訂可能率，作為旺季時客戶的訂房優先順序的判定依據之一。
- 搭配預測來客數保留後的客房數量、預測客戶忠誠度分數以及預測客戶退訂的結果後，提供客戶免費或優惠房型升等，超值房型推薦或同等價位不同房型推薦。



## 2. 專題分項介紹



# Questions & Challenges

1. 現有資料欄位是否都是必須的？
2. 現有資料內容包含空值！
3. Reviews資料中的欄位是否有信息可以幫助預測？
4. 公開資料中並沒有經營數據（如：客戶名稱、住房率...）





### 3. 應用與結論



# 專題分項介紹

- 本組設定了 4 個問題進行研究：
- 來客數預測
- 客戶忠誠度預測
- 預訂取消預測
- 推薦房型預測
- 組員們使用此數據集，導入機器學習技術對不同算法的性能進行基準測試，以解決相同類型的問題（分類，分割或其他問題）後找出關鍵因子。

# 來客數預測內容大綱

- 1.1 資料處理
- 1.2 來客數預測 Model
- 1.3 實際應用在來客數預測APP



# 來客數預測內容大綱



- 1.1 資料處理
  - 先將Dataset中的City與Resort分離
  - 再將cancel紀錄刪除並取出Customer\_type為散客的data
  - 處理data的日期欄位
  - 處理欄位的空值
  - 將各筆資料的guest加總
  - 將guest的房型做單熱編碼
  - 以處理過的日期groupby所有data取得以日期為列的table，並sum每日旅客人數，預定房型
  - 檢視並探索新的dataset



# 來客數預測內容大綱



## 資料處理流程(City.ver)

```
df_city = df_copy[(df_copy.hotel == 'City Hotel')&((df_copy.customer_type == 'Transient')|(df_copy.customer_type == 'Transient-Party'))&(df_copy.is_canceled == 0)]
df_city['arrival_date'] = df_city[['arrival_date_year','arrival_date_month','arrival_date_day_of_month']].astype(str).agg('-'.join,axis = 1)
df_city[['reservation_status_date','arrival_date']] = df_city[['reservation_status_date','arrival_date']].apply(pd.to_datetime)
df_city['children']=df_city['children'].fillna(0)
df_city['children']=df_city['children'].astype(int)
df_city['total_guest'] = df_city['adults'] + df_city['children'] + df_city['babies']
df_city['arrival_day_of_week'] = df_city['arrival_date'].dt.day_name()
df_city['reservation_day_of_week'] = df_city['reservation_status_date'].dt.day_name()
df_city[['A_room','B_room','C_room','D_room','E_room','F_room','G_room','K_room']] = pd.get_dummies(df_city.assigned_room_type)
df_sel_city =
df_city.groupby('arrival_date')[['adults','children','babies','total_guest','A_room','B_room','C_room','D_room','E_room','F_room','G_room','K_room']].sum()
df_sel_city['day_name'] = df_sel_city.index.day_name()
df_sel_city =
df_sel_city[['day_name','total_guest','adults','children','babies','A_room','B_room','C_room','D_room','E_room','F_room','G_room','K_room']]
```

# 來客數預測內容大綱



- 1.2 來客數預測Model
  - 檢視dataset的離群值並了解各欄位屬性
  - 確認預測目標
  - 以各個Model以及所有欄位先run過
  - 比較各個Model的準確度
  - 挑選合適的Model後微調參數及試圖刪減欄位

# 來客數預測內容大綱



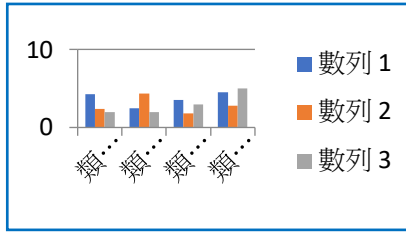
- 1.3 實際應用在來客數預測APP
  - 以現有訂單紀錄，推估該日佔房率
  - 若預測該日將佔房率超過75%時給予通知



# 專題分項介紹

- 本組設定了 4 個問題進行研究：
  - 來客數預測
  - 客戶忠誠度預測
  - 預訂取消預測
  - 推薦房型預測
- 組員們使用此數據集，導入機器學習技術對不同算法的性能進行基準測試，以解決相同類型的問題（分類，分割或其他問題）後找出關鍵因子。

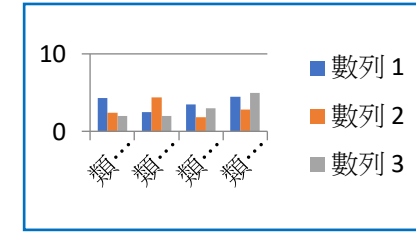




來客分析淡/旺季

| is_repeated_guest | children | babies | deposit_type | days_in_waiting_list | ADR | 取消預訂率(%) | 預備金(美金) |
|-------------------|----------|--------|--------------|----------------------|-----|----------|---------|
| 0                 | 0        | 0      | No deposit   | 0                    | 88  | 0        | 40      |
| 1                 | 0        | 0      | No deposit   | 0                    | 92  | 0        | 40      |
| 1                 | 1        | 0      | No deposit   | 0                    | 80  | 0        | 35      |
| 1                 | 0        | 1      | No deposit   | 0                    | 88  | 0        | 35      |
| 1                 | 1        | 1      | No deposit   | 0                    | 88  | 0        | 35      |
| 1                 | 0        | 0      | No deposit   | 1                    | 87  | 0        | 35      |
| 1                 | 1        | 0      | No deposit   | 1                    | 86  | 0        | 35      |
| 1                 | 1        | 1      | No deposit   | 1                    | 85  | 0        | 35      |
| 1                 | 0        | 1      | No deposit   | 2                    | 83  | 0        | 30      |
| 1                 | 1        | 0      | No deposit   | 2                    | 82  | 0        | 30      |
| 1                 | 0        | 1      | No deposit   | 2                    | 82  | 0        | 30      |
| 1                 | 1        | 1      | No deposit   | 2                    | 80  | 0        | 25      |
| 0                 | 0        | 0      | No deposit   | 0                    | 79  | 0        | 25      |
| 0                 | 0        | 0      | No deposit   | 1                    | 78  | 0        | 25      |
| 0                 | 0        | 0      | No deposit   | 1                    | 77  | 0        | 25      |
| 0                 | 0        | 0      | No deposit   | 2                    | 76  | 0        | 25      |
| 0                 | 0        | 0      | No deposit   | 3                    | 76  | 0        | 20      |
| 0                 | 0        | 0      | Refundable   | 1                    | 74  | 0        | 20      |
| 0                 | 0        | 0      | Refundable   | 2                    | 74  | 0        | 20      |

維護忠誠度量表



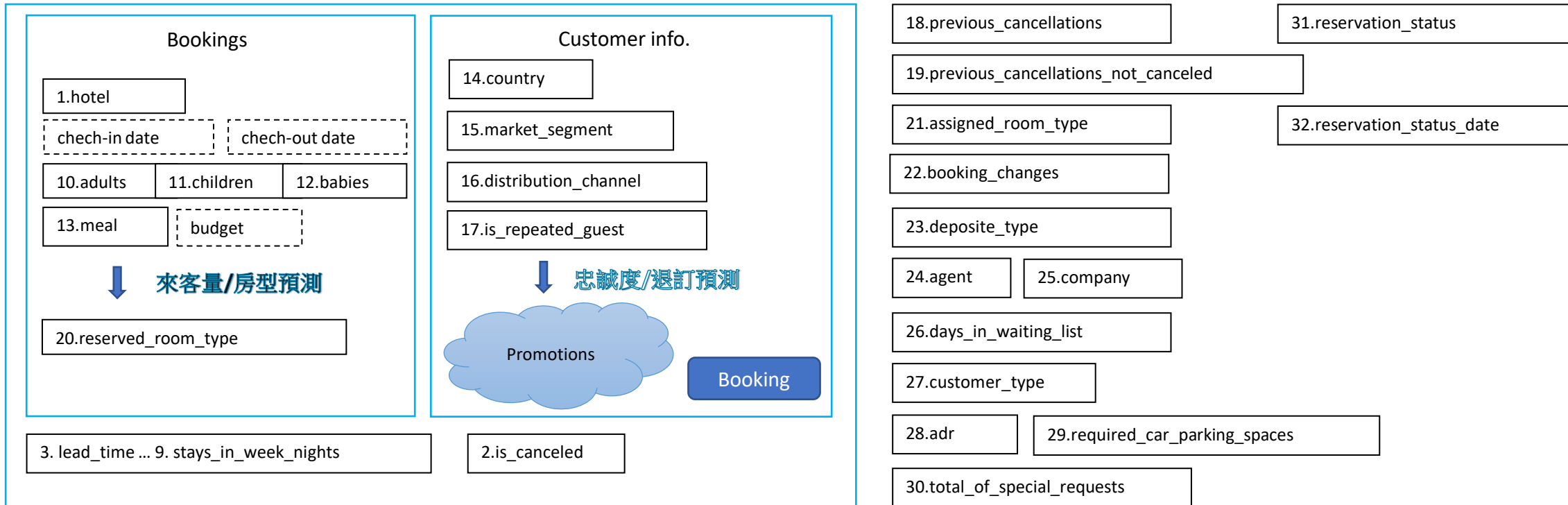
退訂分析

Initial booking

revise/confirm

退訂預測TBC

check-out/ cancel



## 維護忠誠度量表

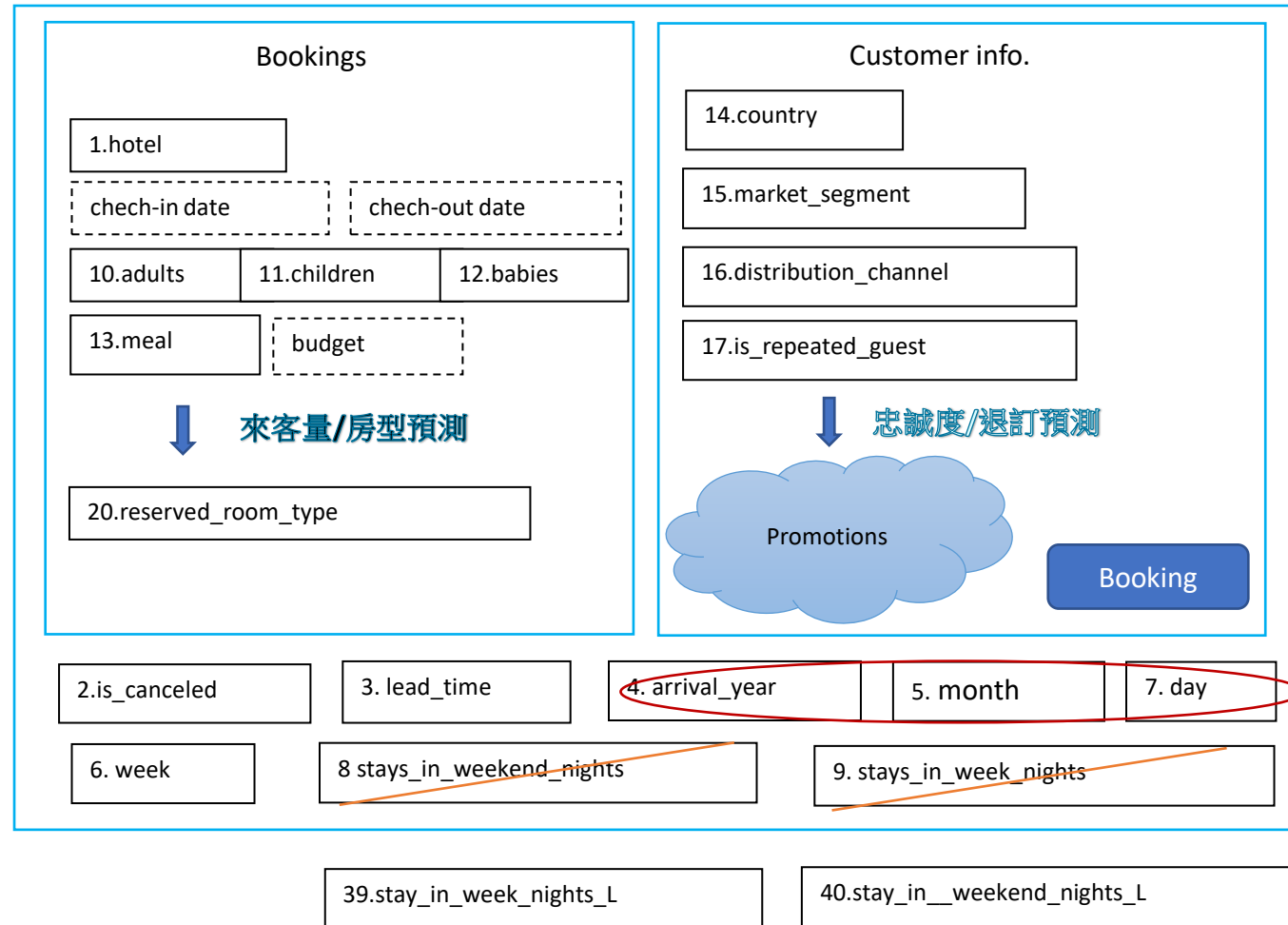


| is_repeated_guest | children | babies | deposit_type | days_in_waiting_list | AI 預測忠誠度分數 | = | 獎勵優惠折扣 |
|-------------------|----------|--------|--------------|----------------------|------------|---|--------|
| 3                 | 0        | 0      | No deposit   | 0                    | 93         |   | 40     |
| 2                 | 0        | 0      | No deposit   | 0                    | 92         |   | 40     |
| 1                 | 0        | 0      | No deposit   | 0                    | 91         |   | 40     |
| 1                 | 1        | 0      | No deposit   | 0                    | 90         |   | 35     |
| 1                 | 0        | 1      | No deposit   | 0                    | 89         |   | 35     |
| 1                 | 1        | 1      | No deposit   | 0                    | 88         |   | 35     |
| 1                 | 0        | 0      | No deposit   | 1                    | 87         |   | 35     |
| 1                 | 1        | 0      | No deposit   | 1                    | 86         |   | 35     |
| 1                 | 0        | 1      | No deposit   | 1                    | 85         |   | 30     |
| 1                 | 1        | 1      | No deposit   | 1                    | 84         |   | 30     |
| 1                 | 0        | 0      | No deposit   | 2                    | 83         |   | 30     |
| 1                 | 1        | 0      | No deposit   | 2                    | 82         |   | 30     |
| 1                 | 0        | 1      | No deposit   | 2                    | 81         |   | 30     |
| 1                 | 1        | 1      | No deposit   | 2                    | 80         |   | 25     |
| 0                 | 0        | 0      | No deposit   | 0                    | 79         |   | 25     |
| 0                 | 0        | 0      | No deposit   | 1                    | 78         |   | 25     |
| 0                 | 0        | 0      | No deposit   | 2                    | 77         |   | 25     |
| 0                 | 0        | 0      | No deposit   | 3                    | 76         |   | 25     |
| 0                 | 0        | 0      | Refundable   | 0                    | 75         |   | 20     |
| 0                 | 0        | 0      | Refundable   | 1                    | 74         |   | 20     |
| 0                 | 0        | 0      | Refundable   | 2                    | 73         |   | 20     |

# 退訂預測



- 資料清洗



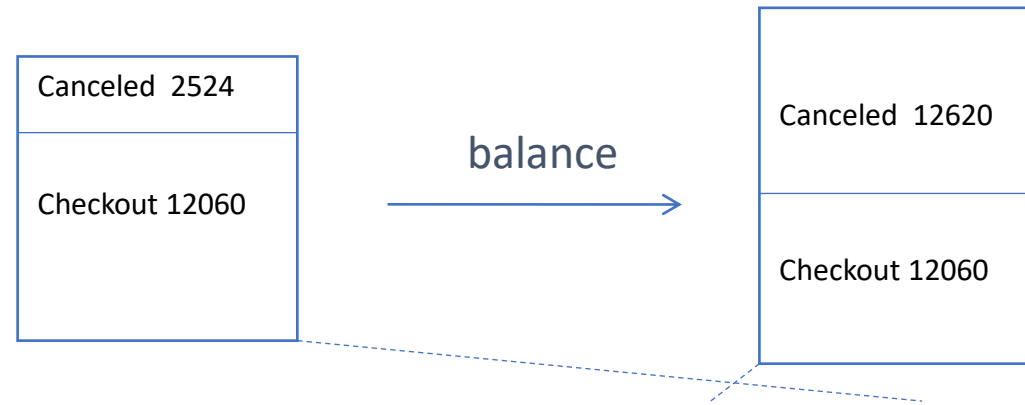
# 退訂預測



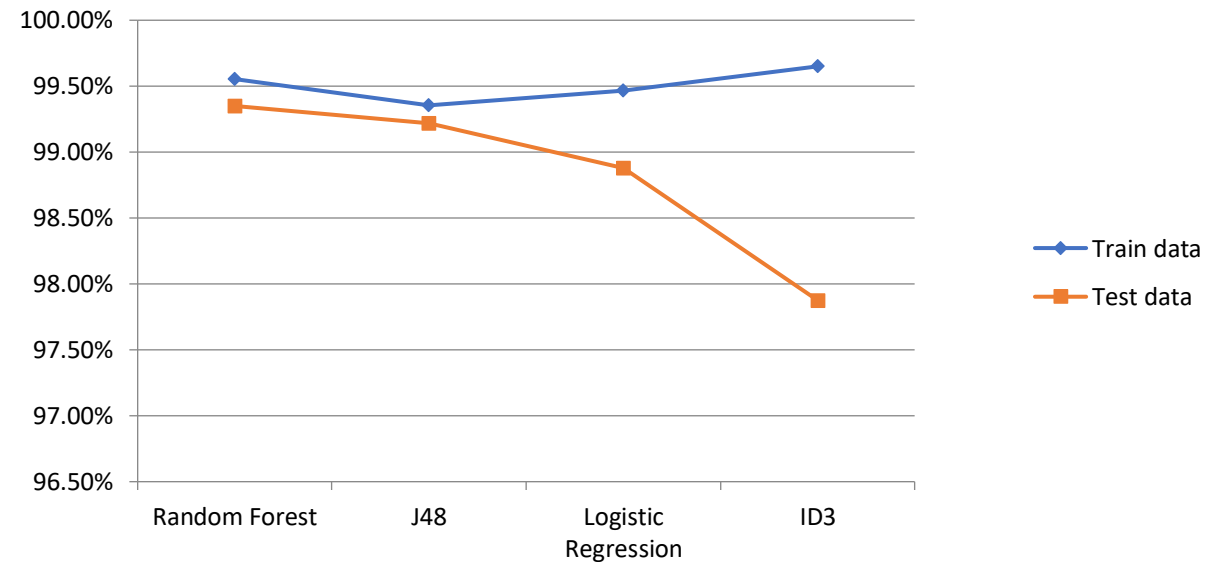
## Machine Learning Model

16. distribution\_channel='Direct'  
3. lead\_time <365  
: 14584筆

- hotel
  - arrival\_date\_month\_num\_L
  - arrival\_date\_day\_of\_month
  - lead\_time
  - stay\_nights\_L
  - stay\_week\_nights\_L
  - stay\_weekend\_nights\_L
  - adults
  - children
  - babies
  - meal
  - reserved\_room\_type
  - is\_repeated\_guest
  - country
  - market\_segment
- is\_canceled



| model               | Train data | Test data |
|---------------------|------------|-----------|
| Logistic Regression | 99.47%     | 98.88%    |
| ID3                 | 99.65%     | 97.87%    |
| J48                 | 99.35%     | 99.22%    |
| Random Forest       | 99.55%     | 99.35%    |





# 退訂預測 TBC



- Machine Learning Model**

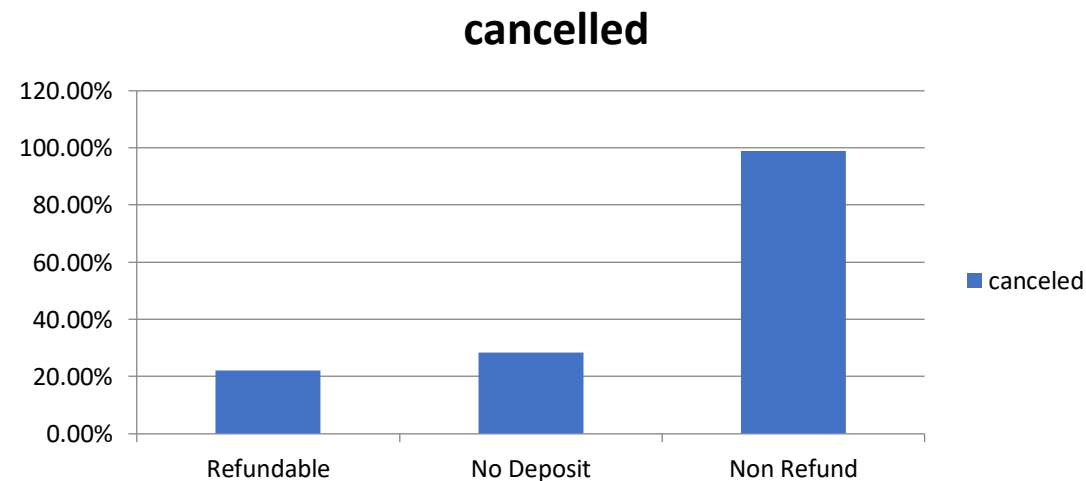
Booking → confirm → checkout/cancel:  
116090筆

增減相關欄位 try diff. model.



- Interesting points : Deposit type: Non-refund v.s. cancellations**

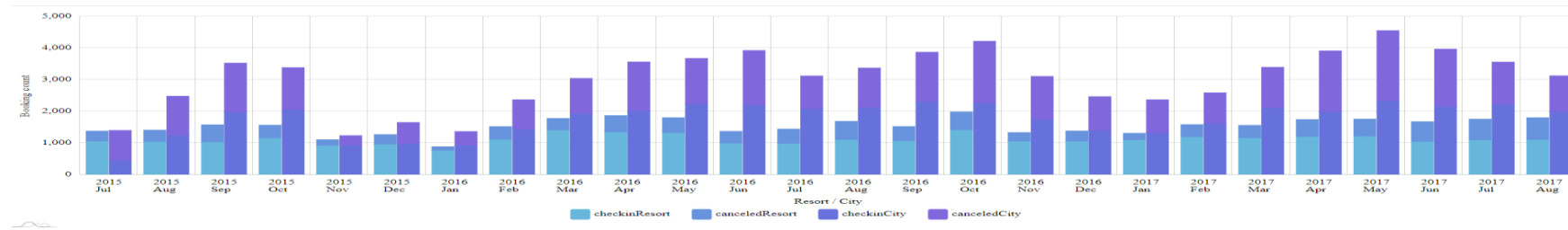
| deposite_type |             |        |              |
|---------------|-------------|--------|--------------|
| Non Refund    | is_cancel=0 | 12772  | 93 0.73%     |
|               | is_cancel=1 | 12629  | 98.88%       |
| No Deposit    | is_cancel=0 | 103208 | 73812 71.52% |
|               | is_cancel=1 | 29396  | 28.48%       |
| Refundable    | is_cancel=0 | 159    | 124 77.99%   |
|               | is_cancel=1 | 35     | 22.01%       |



## 退訂預測 TBC



- find possible cancelations



- 整套 Migrate 到 Hadoop



# 專題分項介紹

- 本組設定了 4 個問題進行研究：
  - 來客數預測
  - 客戶忠誠度預測
  - 預訂取消預測
  - 推薦房型預測
- 組員們使用此數據集，導入機器學習技術對不同算法的性能進行基準測試，以解決相同類型的問題（分類，分割或其他問題）後找出關鍵因子。

# 推薦房型預測



- Hotel 主要分成 City Hotel 及 Resort Hotel
- Hotel 房型主要有 A,B,C,D,E,F,G,H 等房型 (Y)
- 主要影響房型變異因子 ( $X_i$ ) 為
  - 地區 (City , Resort )
  - 住房人數 (成人 , 兒童 , 嬰孩)
  - 房型預算\$
- 最後依照客戶的預算, 入住人數, 訂房網站給予房型建議



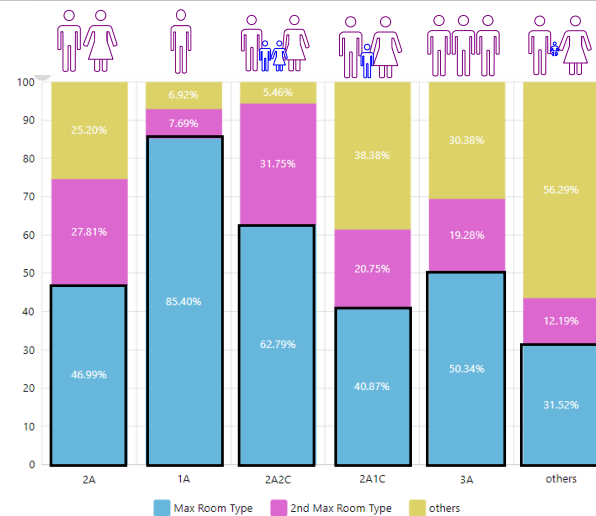
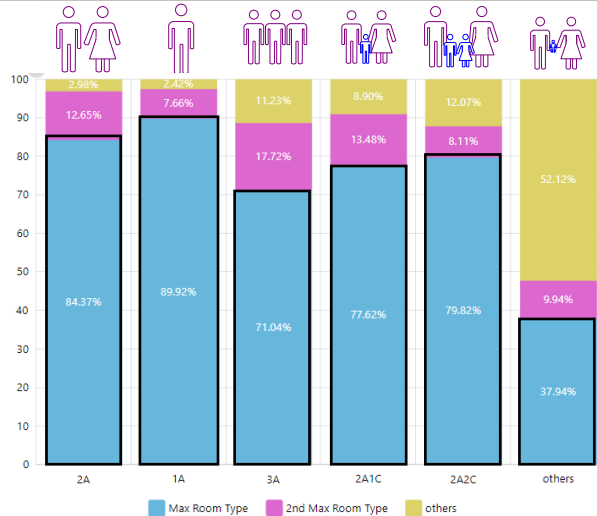




# 入住人數與房型分類統計

| City  | All   | 2A    | 1A    | 3A    | 2A1C  | 2A2C  | 2A1b  | others | Resort | All   | 2A    | 1A    | 2A2C  | 2A1C  | 3A    | 2A1b  | others |
|-------|-------|-------|-------|-------|-------|-------|-------|--------|--------|-------|-------|-------|-------|-------|-------|-------|--------|
| A     | 62595 | 45160 | 13995 | 822   | 2067  | 22    | 206   | 206    | A      | 23399 | 16036 | 5989  | 10    | 526   | 517   | 238   | 238    |
| B     | 1115  | 729   | 33    | 5     | 30    | 56    | 8     | 8      | B      | 3     | 3     | 0     | 0     | 0     | 0     | 0     | 0      |
| C     | 14    | 7     | 5     | 0     | 0     | 1     | 0     | 0      | C      | 918   | 160   | 13    | 494   | 96    | 47    | 13    | 13     |
| D     | 11768 | 6770  | 1192  | 3295  | 359   | 6     | 54    | 54     | D      | 7433  | 6294  | 539   | 5     | 267   | 198   | 92    | 92     |
| E     | 1553  | 601   | 270   | 391   | 73    | 135   | 6     | 6      | E      | 4982  | 4285  | 348   | 1     | 165   | 108   | 56    | 56     |
| F     | 1791  | 156   | 33    | 68    | 111   | 1329  | 9     | 9      | F      | 1106  | 893   | 72    | 1     | 66    | 50    | 13    | 13     |
| G     | 484   | 100   | 36    | 57    | 23    | 116   | 0     | 0      | G      | 1610  | 311   | 46    | 977   | 122   | 64    | 13    | 13     |
| H     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0      | H      | 601   | 48    | 5     | 68    | 45    | 43    | 0     | 0      |
| L     | 0     | 0     | 0     | 0     | 0     | 0     | 0     | 0      | L      | 6     | 4     | 1     | 0     | 0     | 0     | 0     | 0      |
| P     | 10    | 0     | 0     | 0     | 0     | 0     | 0     | 0      | P      | 2     | 0     | 0     | 0     | 0     | 0     | 0     | 0      |
| Max比例 |       | 84.4% | 89.9% | 71.0% | 77.6% | 79.8% | 72.8% | 72.8%  | Max比例  |       | 57.2% | 85.4% | 62.8% | 40.9% | 50.3% | 56.0% | 56.0%  |

- X軸是住房人數及地區，Y軸是房型，主要選擇為A房型及D房型。但是2A2C兩大兩小則選擇其他房型。
- 下圖藍色標示最大比例房型，粉紅色為第二大比例的房型，土黃色為其他比例的房型。
- Resort Hotel 兩人以上選擇性較多，藍色區塊比例明顯變少。





## 4.2.1 由原始資料推斷房型數量及價格範圍



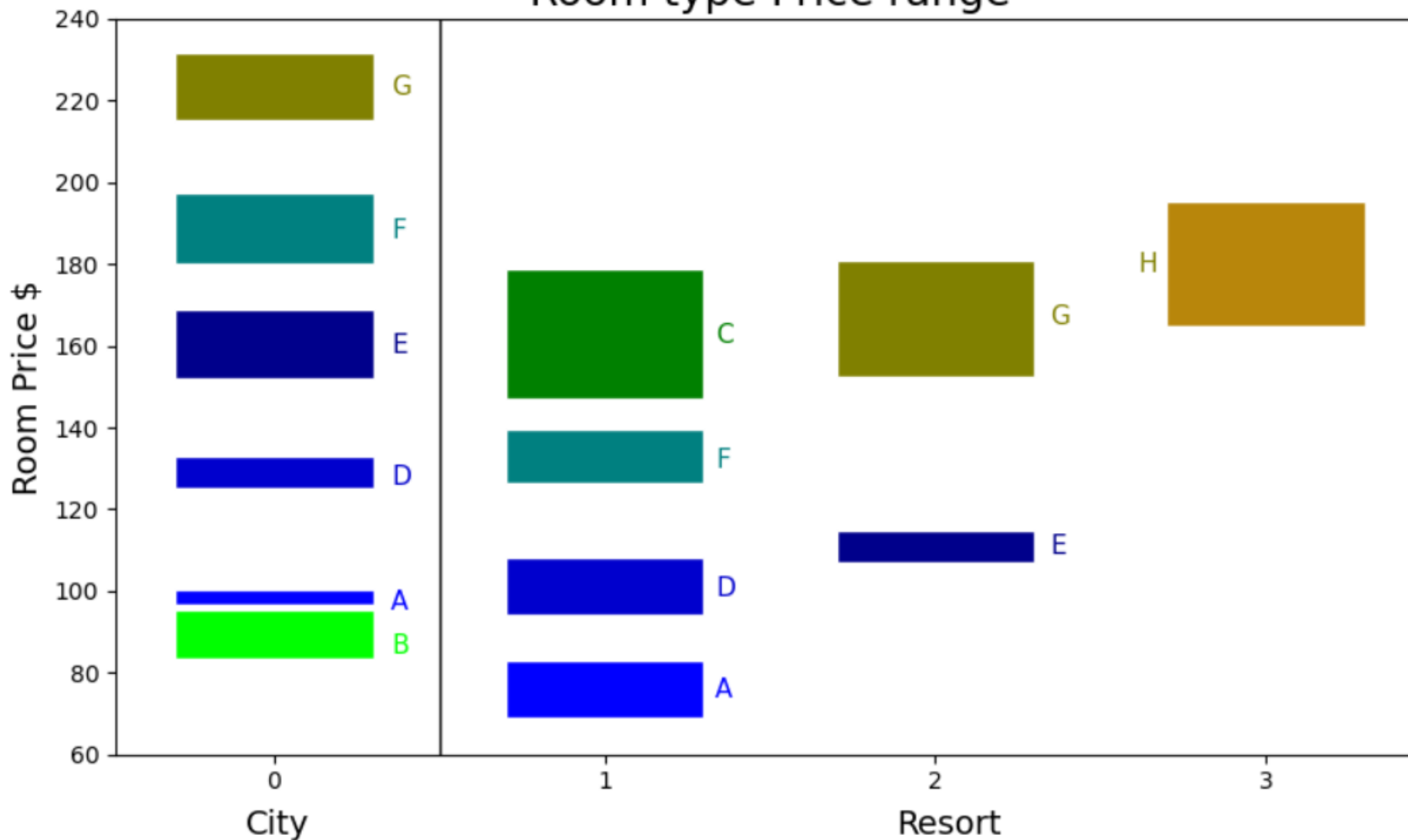
| 區域     | 房間類型 | 價格範圍  |       | 房間數 |
|--------|------|-------|-------|-----|
| City   | B    | 83.3  | 95.6  | 14  |
|        | A    | 96.7  | 100.3 | 217 |
|        | D    | 125.3 | 133.2 | 64  |
|        | E    | 152.2 | 169.1 | 13  |
|        | F    | 180.1 | 197.6 | 9   |
|        | G    | 215   | 231.6 | 5   |
| Resort | A    | 68.8  | 83.0  | 142 |
|        | D    | 94.0  | 108.2 | 64  |
|        | E    | 106.9 | 114.9 | 38  |
|        | F    | 126.3 | 139.8 | 12  |
|        | C    | 146.9 | 179.0 | 14  |
|        | G    | 152.6 | 181.0 | 10  |
|        | H    | 164.8 | 195.2 | 5   |

- 價格範圍是每種房型禮拜一到禮拜日平均價格7個數字最小值和最大值。
- 房間數是2015.7.1~2016.8.31每日各房型使用量的最大值。因此實際房間數可能比預估值還要高。

# 各房型價格



Room type Price range

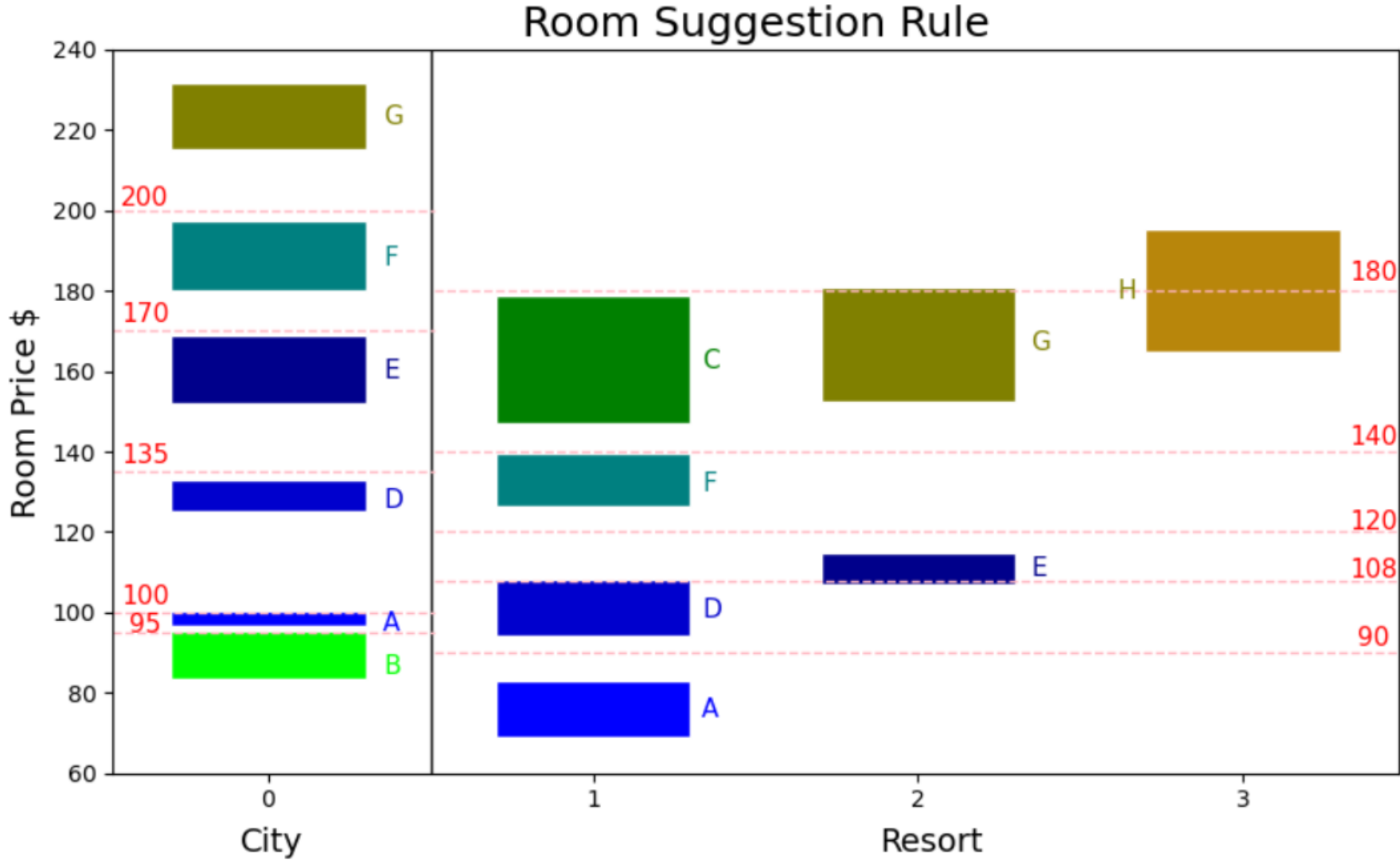


相同房型City Hotel價格會比Resort Hotel還高。Resort Hotel 房型選擇比較多因為Resort Hotel 可以提供更便宜的價格或者更好的房型作選擇。  
 City A => Resort A,D,E  
 City D => Resort D,E,F  
 City E => Resort C,E,G  
 Resort選擇性變多

# 4.2.3 設定房價範圍供客戶選擇



房價





# 4.2.4 依照來客人數及消費金額推薦房型

| 價格       | City   |     |    |      |      |      |        | 價格       | Resort |     |      |      |     |      |        |
|----------|--------|-----|----|------|------|------|--------|----------|--------|-----|------|------|-----|------|--------|
|          | 2A     | 1A  | 3A | 2A1C | 2A2C | 2A1b | others |          | 2A     | 1A  | 2A2C | 2A1C | 3A  | 2A1b | others |
| ~95      | B      |     |    |      |      |      |        | ~90      | A      | A   |      | A    | A   | A    | A      |
| 95~100   | A      | A   | A  | A    |      | A    | A      | 90~108   | D      | D   |      | D    | D   | D    | D      |
| 100~135  | D      | D   | D  | D    |      | D    | D      | 108~120  | E      | E   |      | E    | E   | E    | E      |
| 135~170  | E      | E   | E  | E    | E    | E    | E      | 120~140  | F      | F   |      | F    | F   | F    | F      |
| 170~200  | F      |     |    |      |      |      |        | 140~180  | CG     |     |      |      |     |      |        |
| 200~     | G      |     |    |      |      |      |        | 180~     | H      |     |      |      |     |      |        |
| Priority | ADBEFG | ADE | DA | ADF  | FEG  | ADF  | AD     | Priority | ADE    | ADE | GCH  | ADG  | ADE | ADE  | AD     |



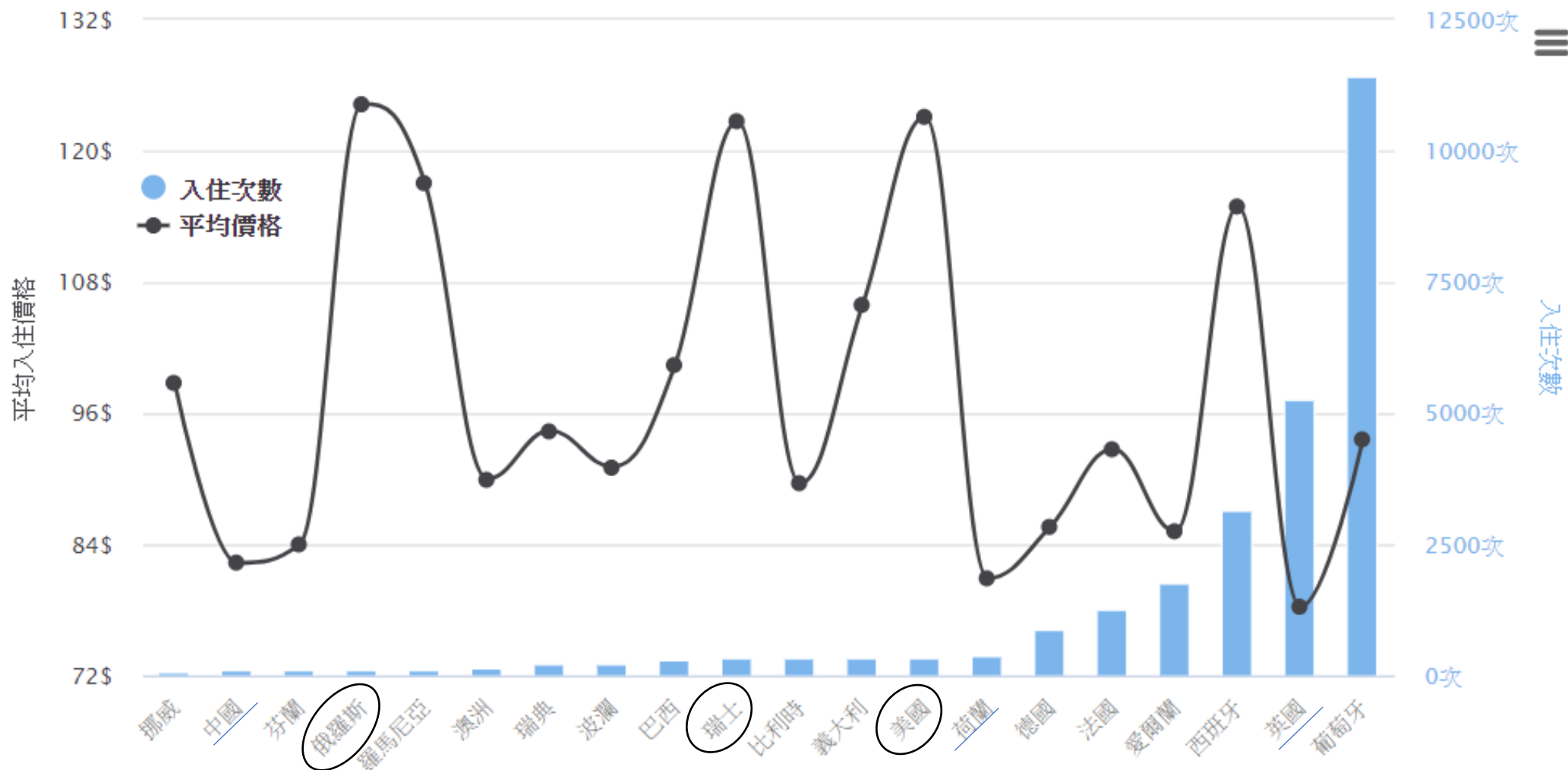
# 4.3.1 資料處理流程,確定決策因子

|              | 地區     | 大人數 | 小孩數 | 嬰兒數 | 供餐 | 國籍 | 顧客型態 | 平均消費金額 | J48   | Xgboost | Random Forest |
|--------------|--------|-----|-----|-----|----|----|------|--------|-------|---------|---------------|
| 1. Raw Dara  | 0      | 0   | 0   | 0   |    |    |      | 0      | 79.43 |         |               |
| 2. Screen離群值 | 0      | 0   | 0   | 0   |    |    |      | 0      | 79.67 |         |               |
|              | City   | 0   | 0   | 0   |    |    |      | 0      | 86.50 |         |               |
|              | Resort | 0   | 0   | 0   |    |    |      | 0      | 65.58 |         |               |
| 3.找尋變因       | 0      | 0   | 0   | 0   | 0  | 0  | 0    | 0      |       | 84.23   | 96.3          |
|              | 0      | 0   | 0   | 0   |    | 0  | 0    | 0      |       | 82.84   | 95.47         |
|              | 0      | 0   | 0   | 0   | 0  |    | 0    | 0      |       | 77.37   | 88.88         |
|              | 0      | 0   | 0   | 0   | 0  |    |      | 0      |       | 76.37   | 86.89         |
|              | 0      | 0   | 0   | 0   |    |    |      | 0      |       | 74.5    | 84.34         |
|              | 0      | 0   | 0   |     | 0  | 0  | 0    | 0      |       | 84.1    | 96.23         |
|              | 0      | 0   | 0   | 0   | 0  | 0  | 0    |        |       | 65.36   | 65.89         |

- 剛開始使用J48分析，拿掉離群值準確率有微幅上升。
- 用Xgboost 及Random Forest來確定決策因子，房型跟地區、來客人數、消費金額、供餐、國籍、顧客型態都有關聯



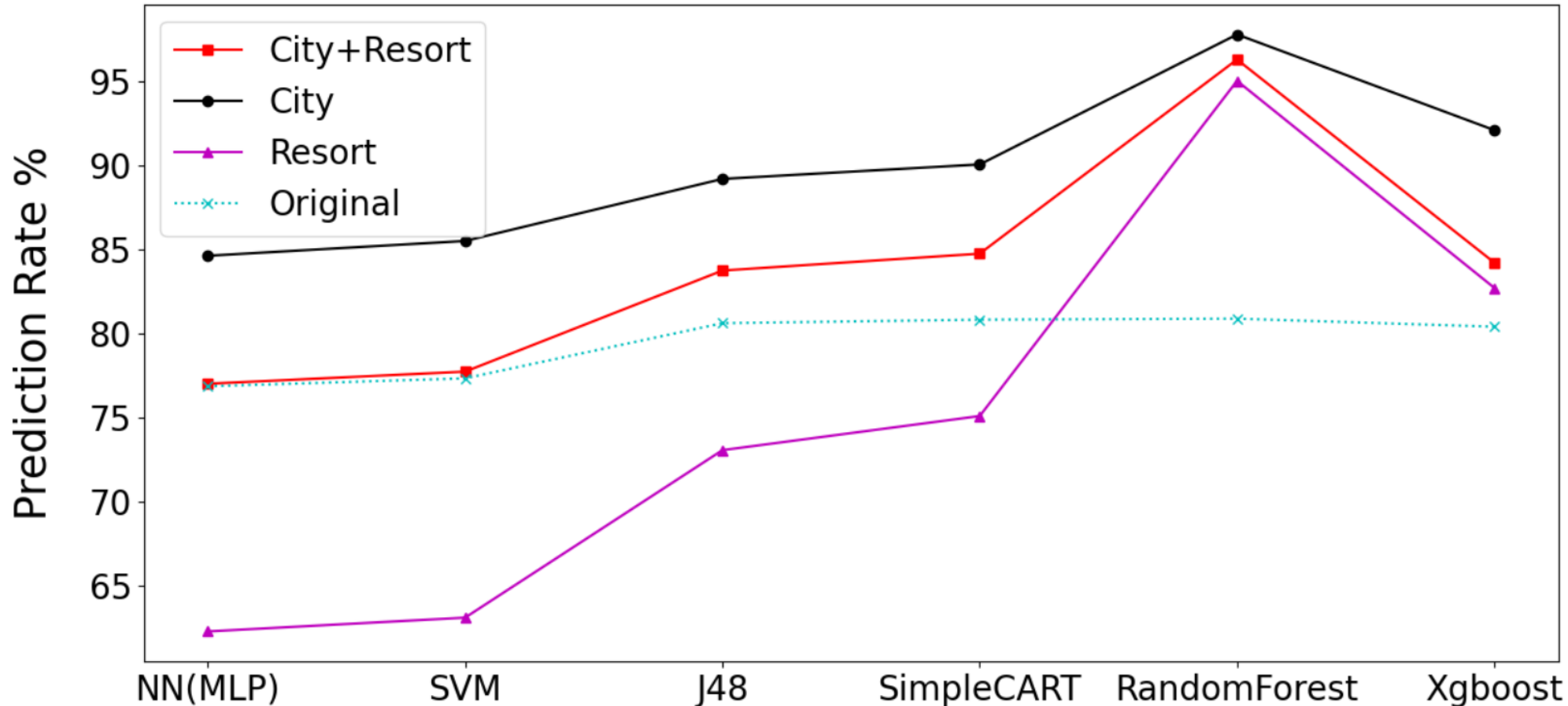
# 主要國家入住次數及平均花費 (Resort 2人)



# 各種方法預測結果



Prediction rate for each method



- 訓練資料處理過效果有變比較好
- City Hotel 比 Resort Hotel 預測值來得好
- MLP、SVM、J48、SaimpleCART 是使用Weka模擬，Random Forest及 Xgboost是使用 Python去模擬。
- 最強準確率的預測方法為 Random Forest