

定期存款認購預測

國泰人壽



鍾佑偉

資料集

銀行行銷資料

<http://archive.ics.uci.edu/ml/datasets/Bank+Marketing#>

程式碼

<https://reurl.cc/arxkY7>



使用語言

Python: Pandas, Numpy, Matplotlib, Seaborn, Plotly, Sklearn, H2O

問題設定

此資料集是來自一家葡萄牙銀行的直接行銷活動數據，直接行銷活動是以電話行銷為基礎，銀行的客服人員至少需要聯繫1次客戶來得知其是否將認購銀行的產品 (定期存款)

預測目標



預測客戶是否認購定期存款

資料集

```
Rows: 41188
Columns: 21
Missing Number Total 0
age          : 78      [56 57 37 40 45]
job          : 12      ['housemaid' 'services' 'admin.' 'blue-collar' 'technician']
marital      : 4       ['married' 'single' 'divorced' 'unknown']
education    : 8       ['basic.4y' 'high.school' 'basic.6y' 'basic.9y' 'professional.course']
default      : 3       ['no' 'unknown' 'yes']
housing      : 3       ['no' 'yes' 'unknown']
loan         : 3       ['no' 'yes' 'unknown']
contact      : 2       ['telephone' 'cellular']
month        : 10      ['may' 'jun' 'jul' 'aug' 'oct']
day_of_week  : 5       ['mon' 'tue' 'wed' 'thu' 'fri']
duration     : 1544    [261 149 226 151 307]
campaign     : 42      [1 2 3 4 5]
pdays       : 27      [999 6 4 3 5]
previous     : 8       [0 1 2 3 4]
poutcome     : 3       ['nonexistent' 'failure' 'success']
emp.var.rate : 10      [ 1.1 1.4 -0.1 -0.2 -1.8]
cons.price.idx : 26     [93.994 94.465 93.918 93.444 93.798]
cons.conf.idx : 26     [-36.4 -41.8 -42.7 -36.1 -40.4]
euribor3m    : 316     [4.857 4.856 4.855 4.859 4.86 ]
nr.employed  : 11      [5191. 5228.1 5195.8 5176.3 5099.1]
y            : 2       ['no' 'yes']
```

電話行銷人員準備撥打時的頁面

- ▶ 聯絡方式: (手機) 0800-036-599
- ▶ 客戶姓名: 國泰人
- ▶ 年齡: 20
- ▶ 工作: 學生
- ▶ 婚姻狀況: 單身
- ▶ 教育程度: 國中畢業
- ▶ 違約紀錄: 無
- ▶ 房子: 有
- ▶ 貸款: 無

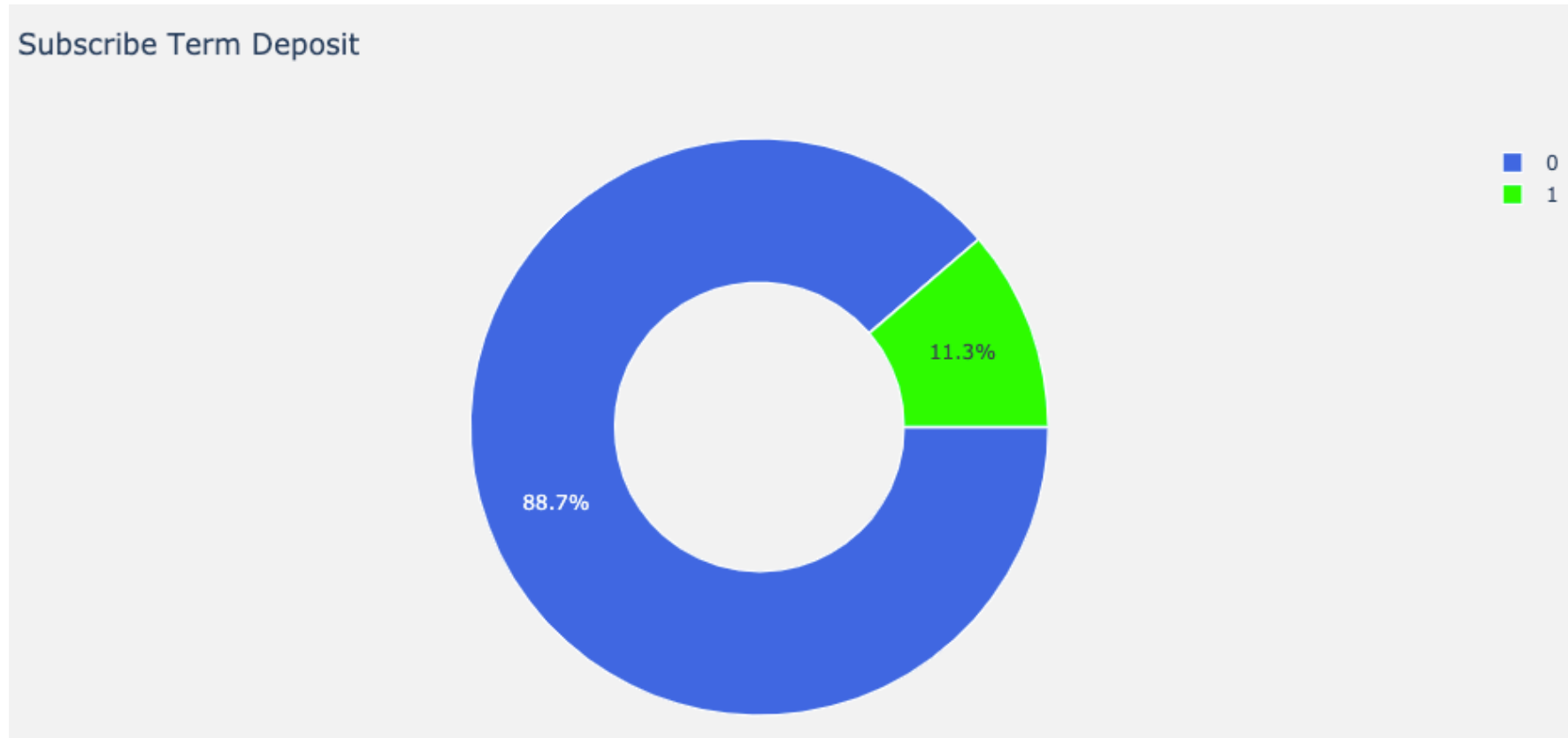
- ▶ 此活動聯絡次數: 2
- ▶ 上次聯絡時間: 6
- ▶ 此活動前聯絡次數: 3
- ▶ 上次活動結果: 成功
- ▶ 電話總長:

預測結果:

?

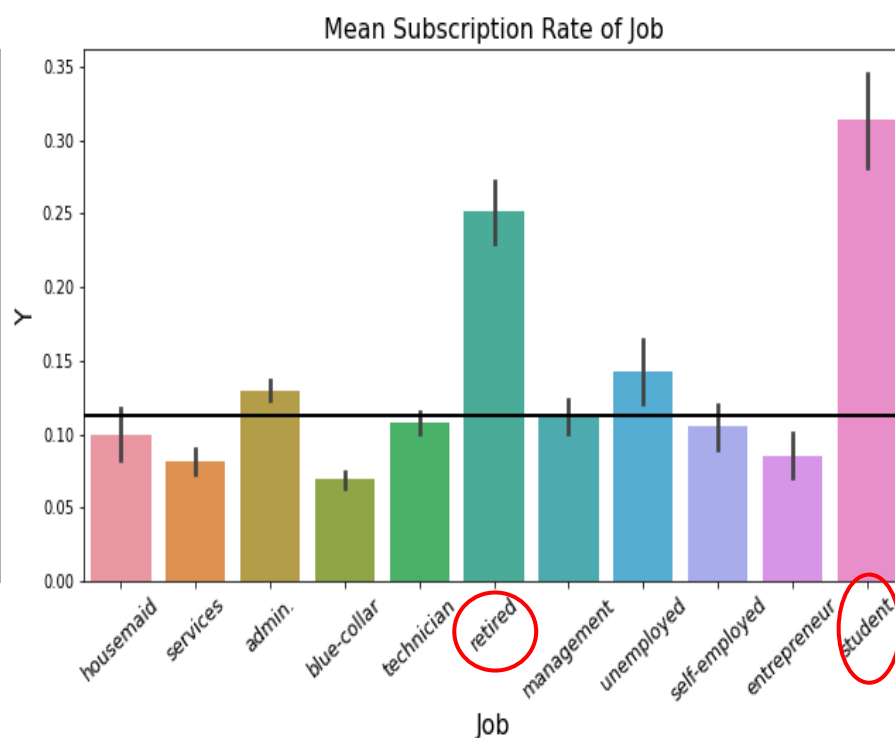
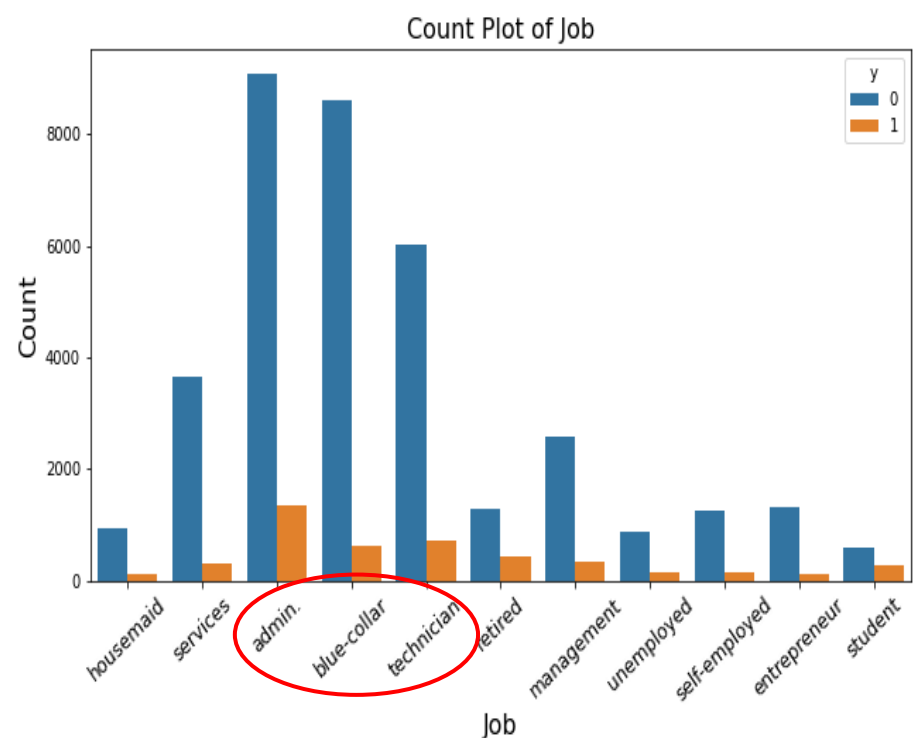
07/04/2019 17:00

目標值

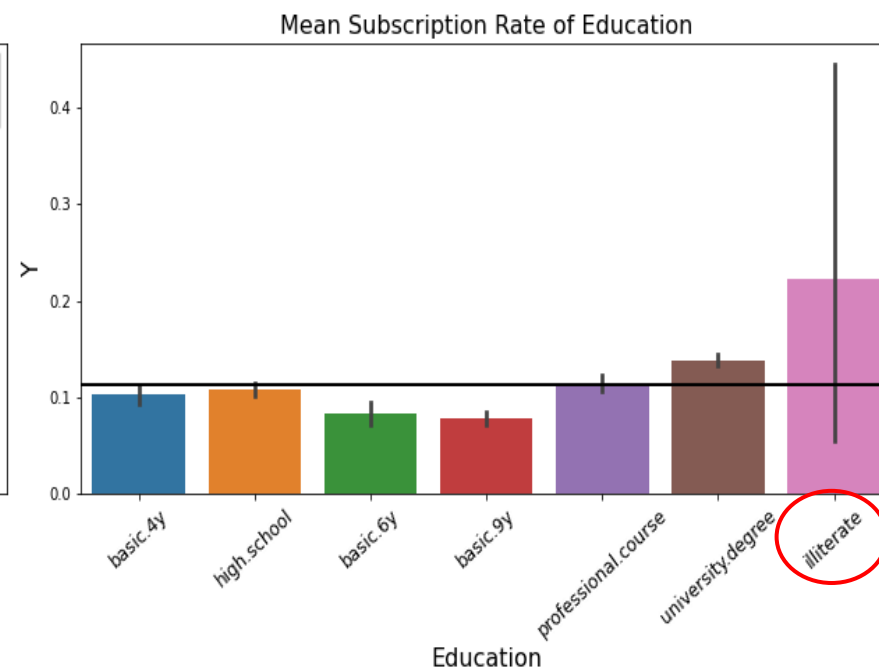
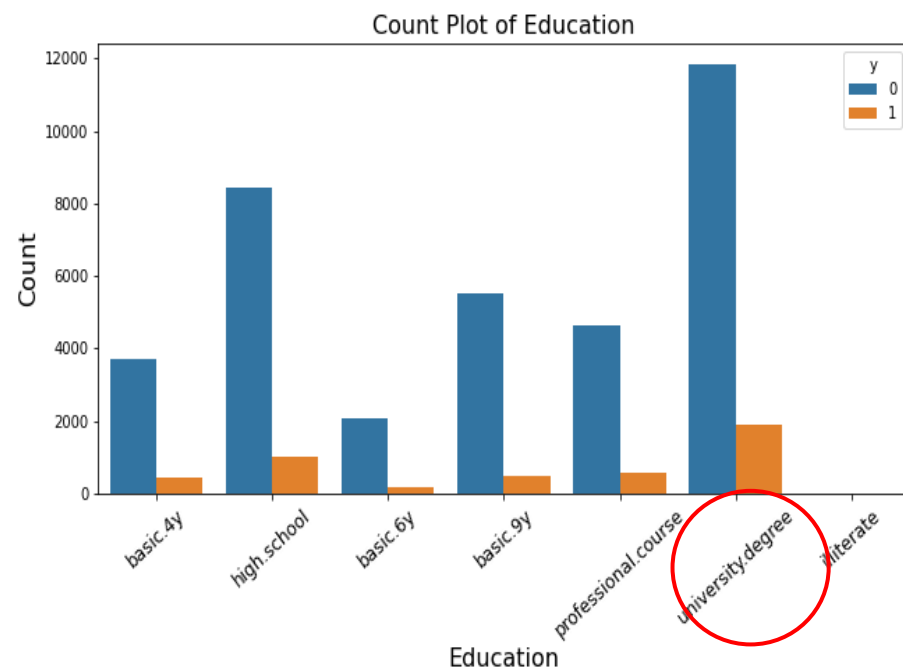


探索性資料分析

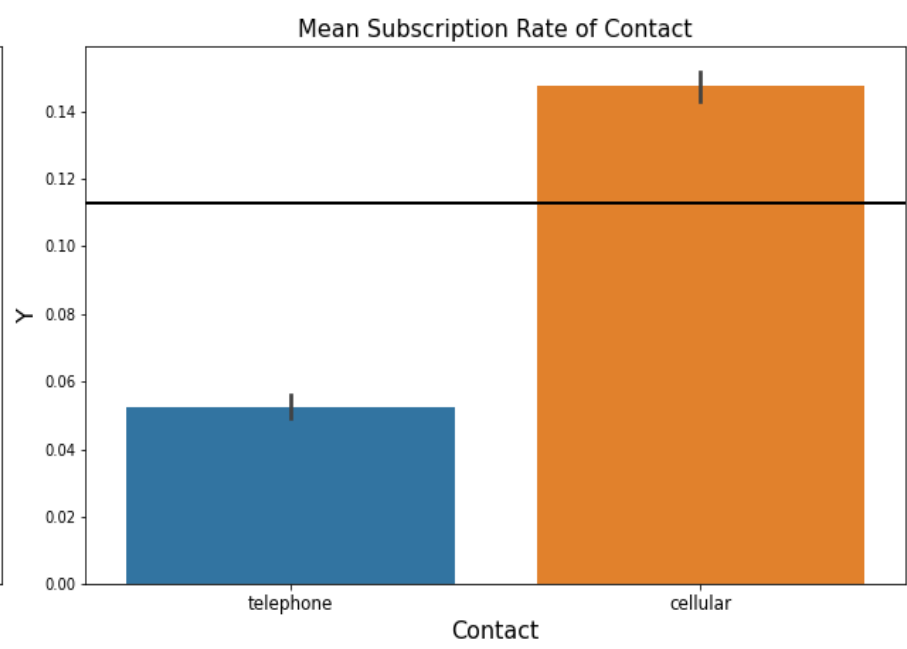
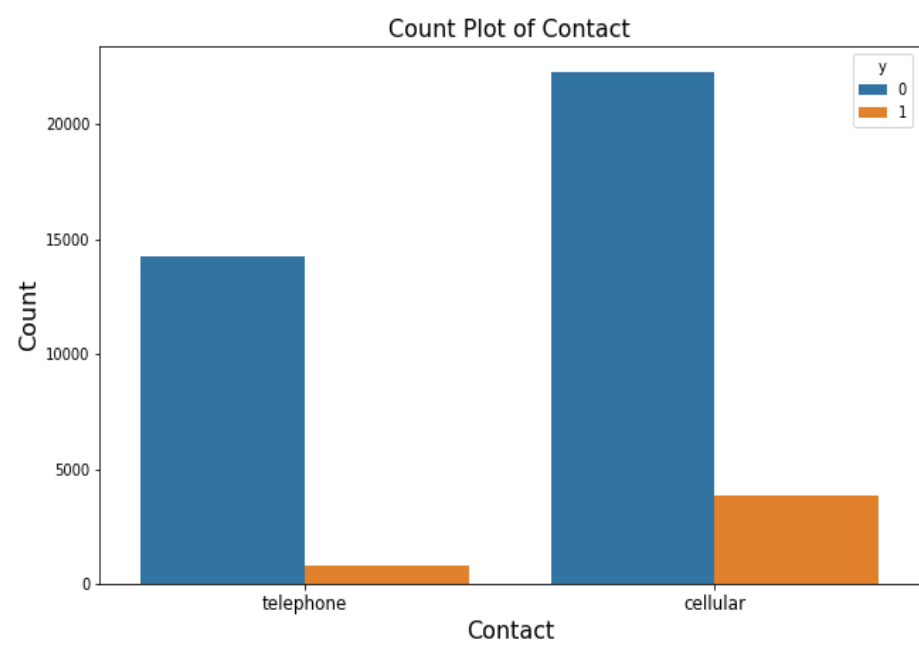
工作型態 (Job)



教育程度 (Education)



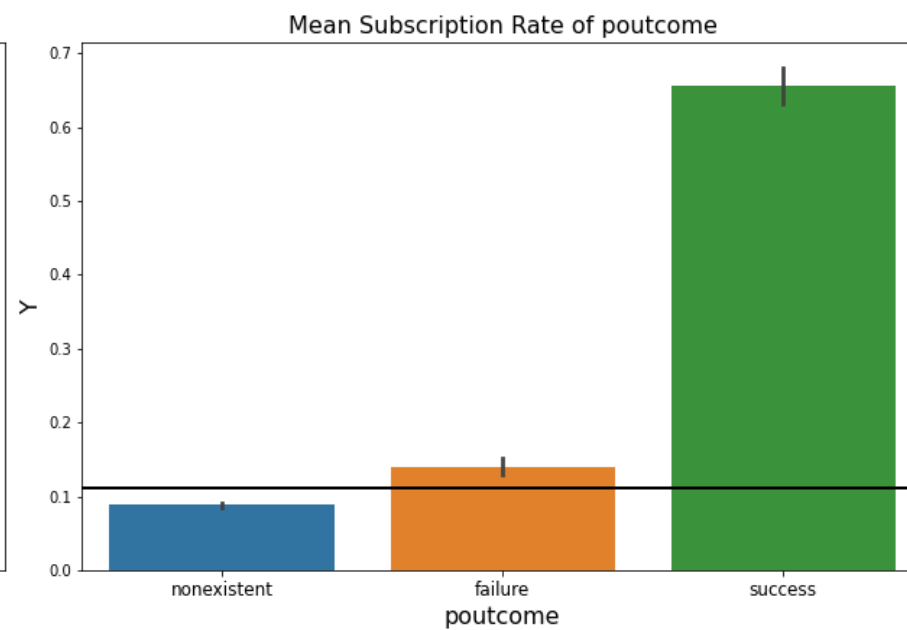
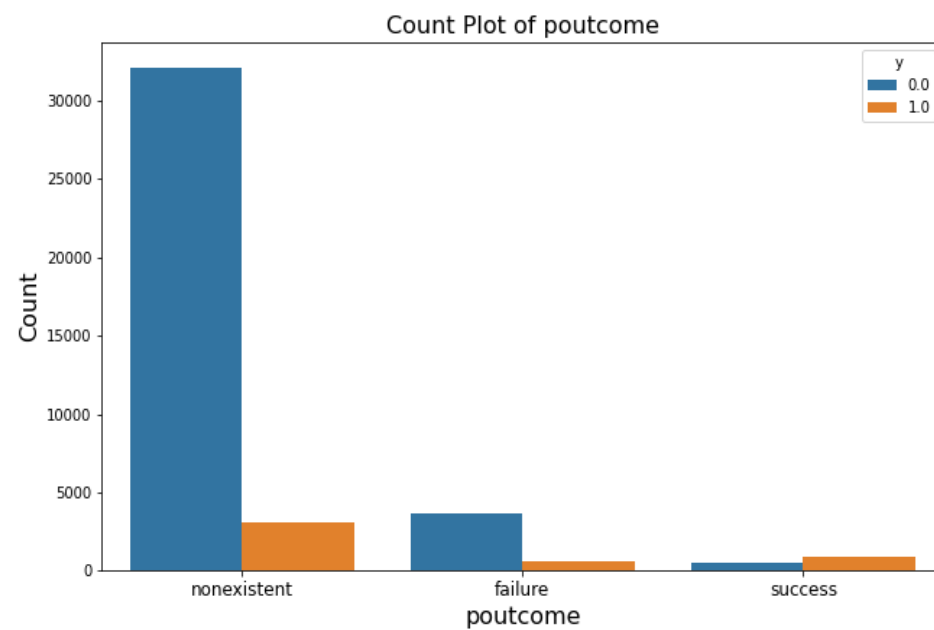
聯絡方式 (Contact)



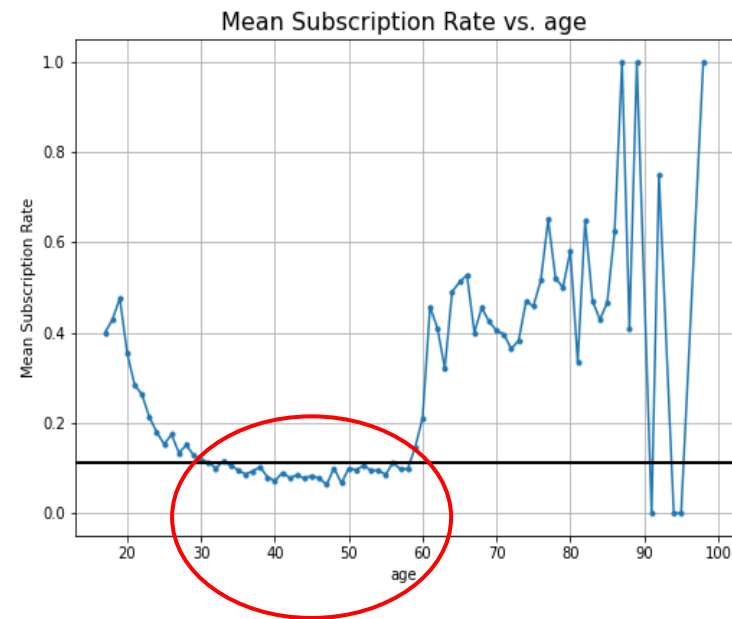
平均認購率

job	contact	
retired	cellular	0.316003
	telephone	0.092025
student	cellular	0.363636
	telephone	0.151961

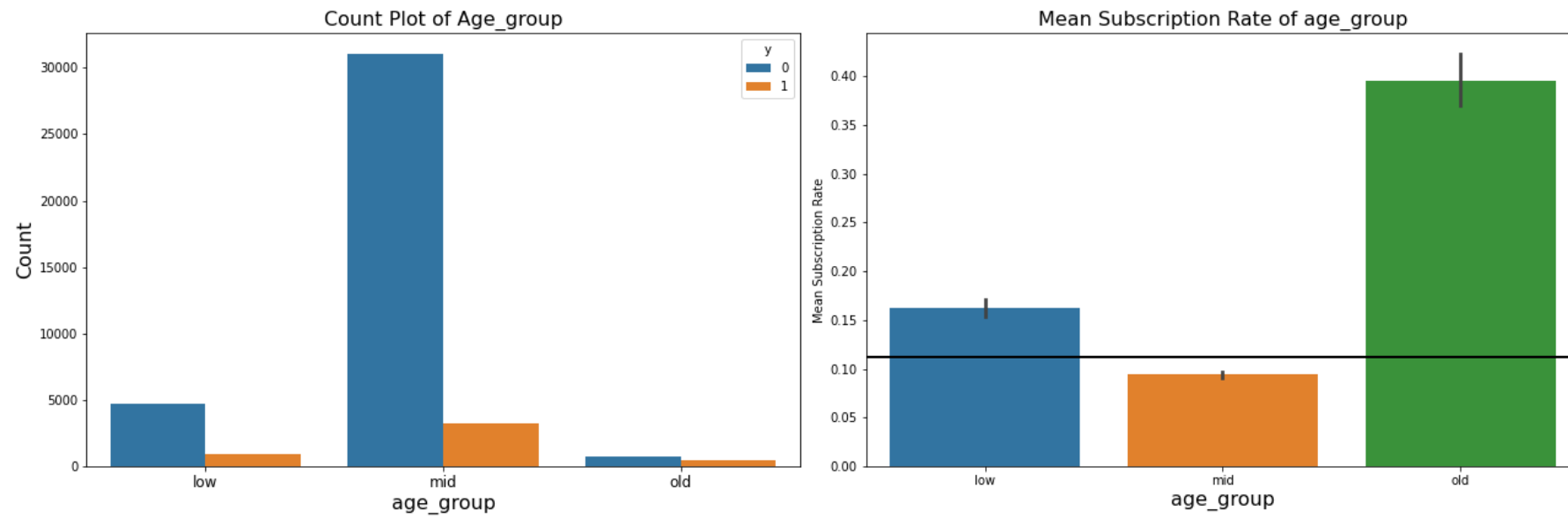
前一次活動的結果 (Poutcome)



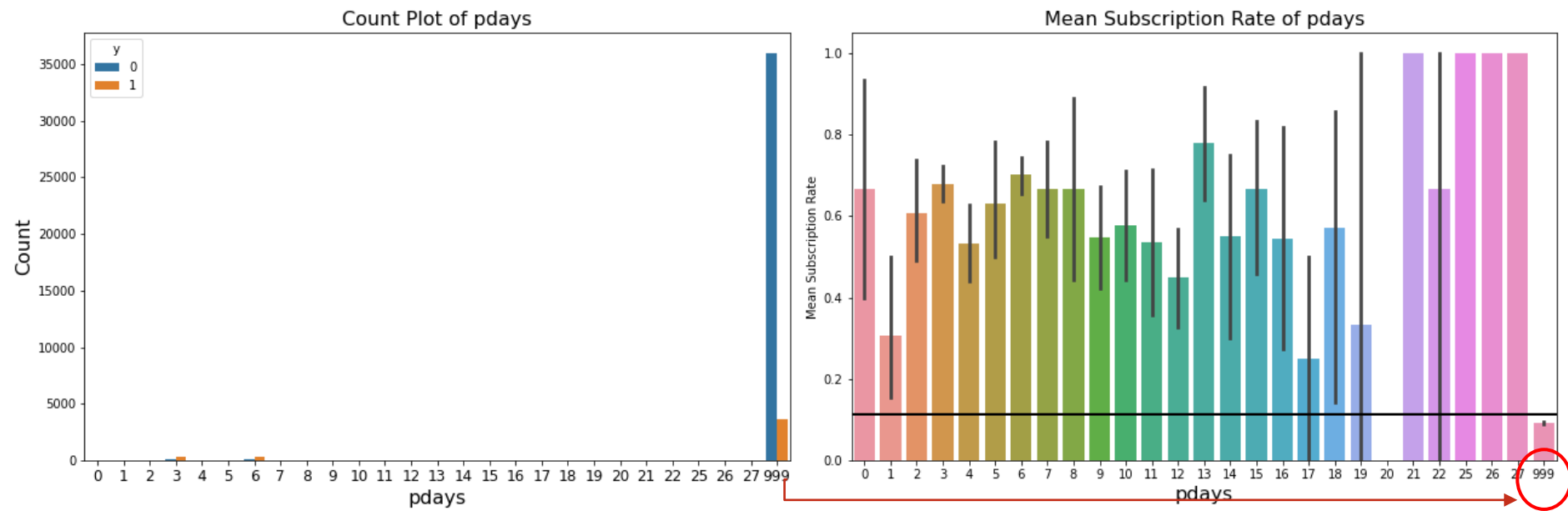
年紀 (Age)



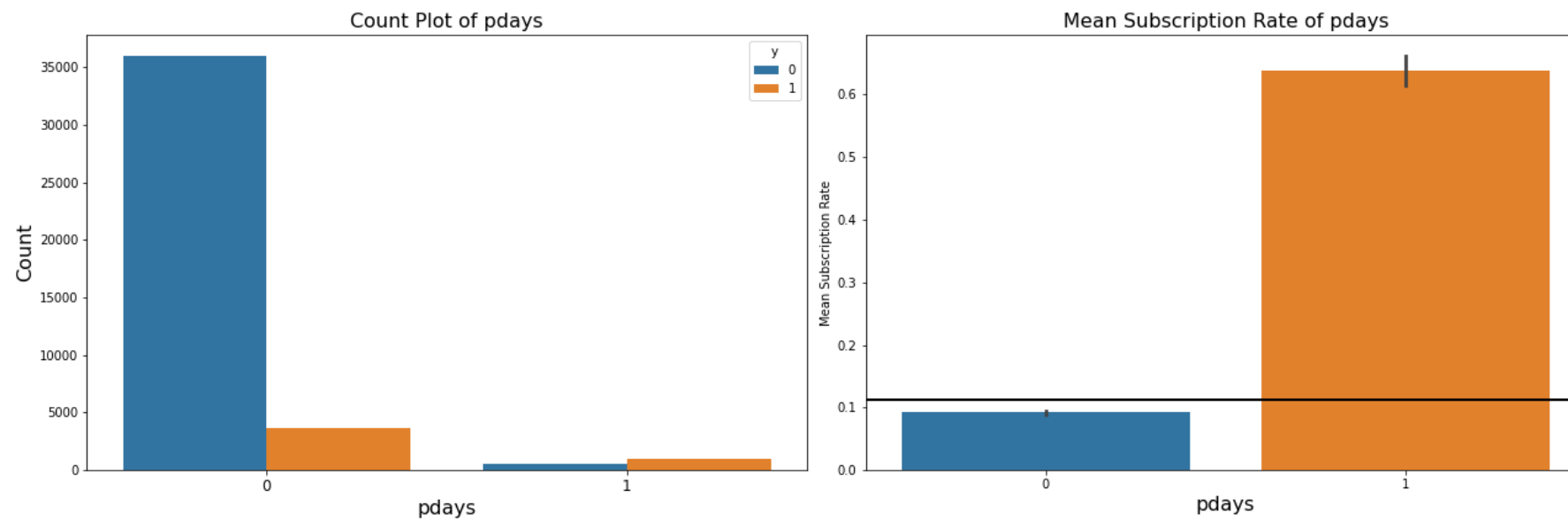
年紀 (Age) - 特徵工程



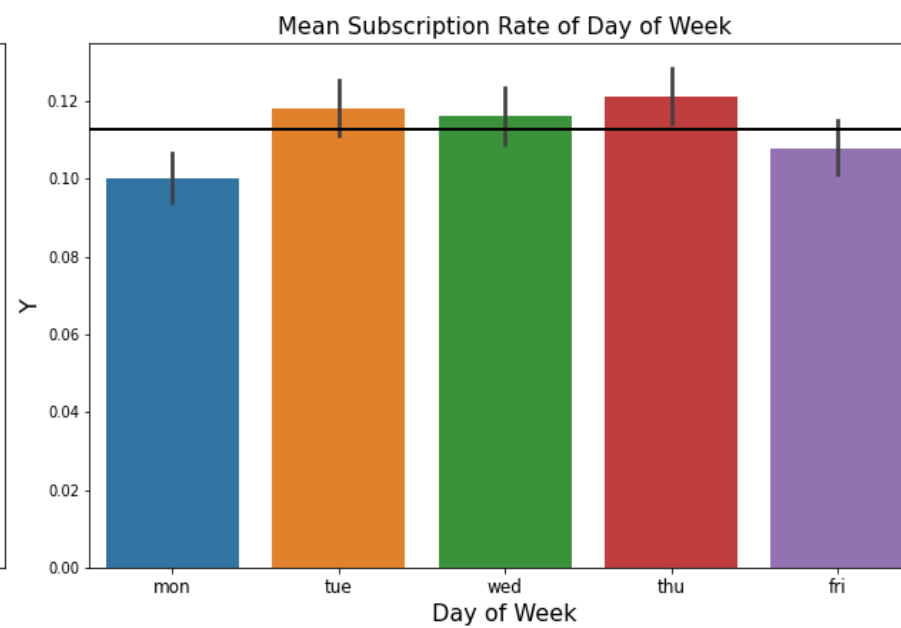
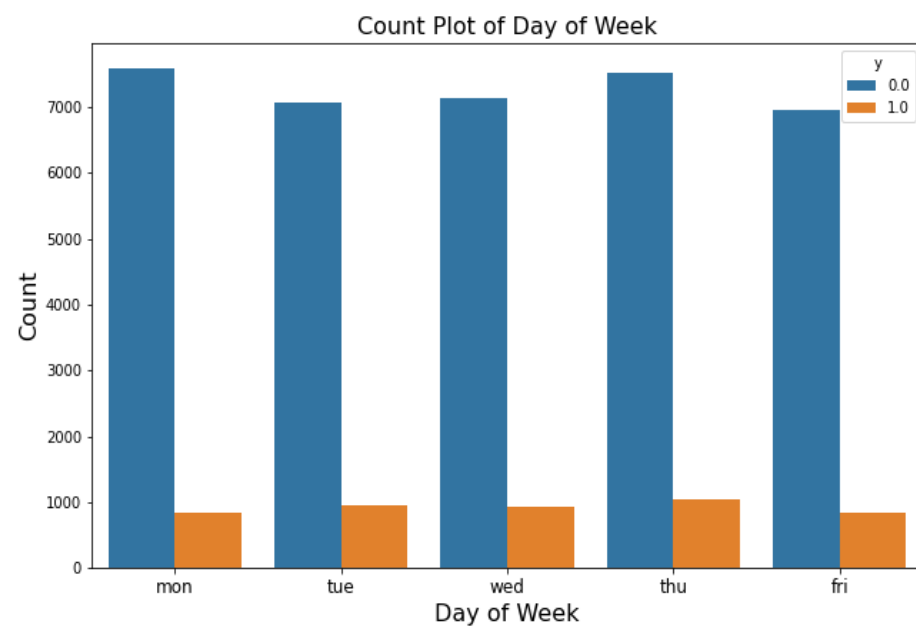
間隔天數(Pdays)



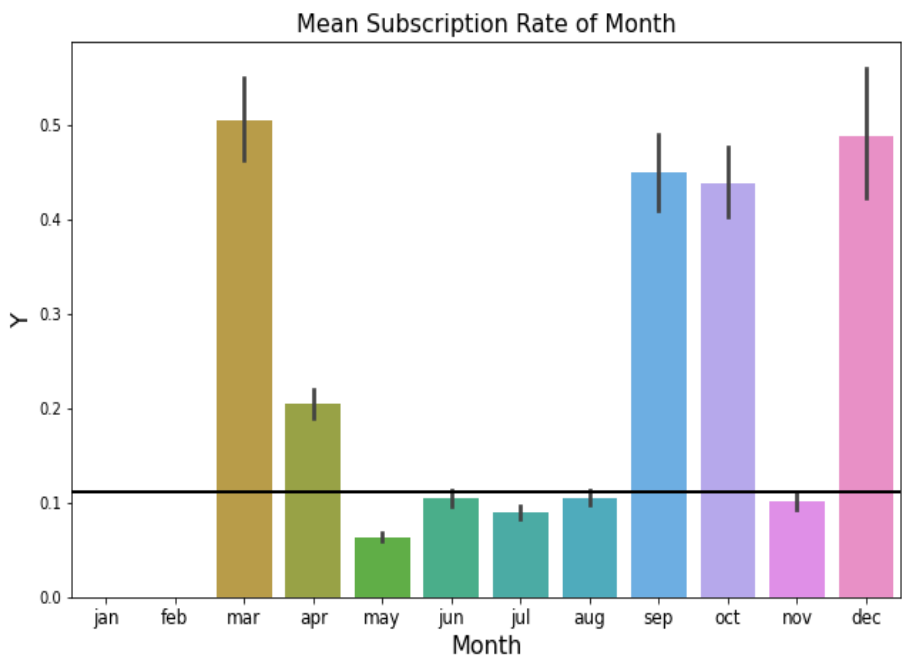
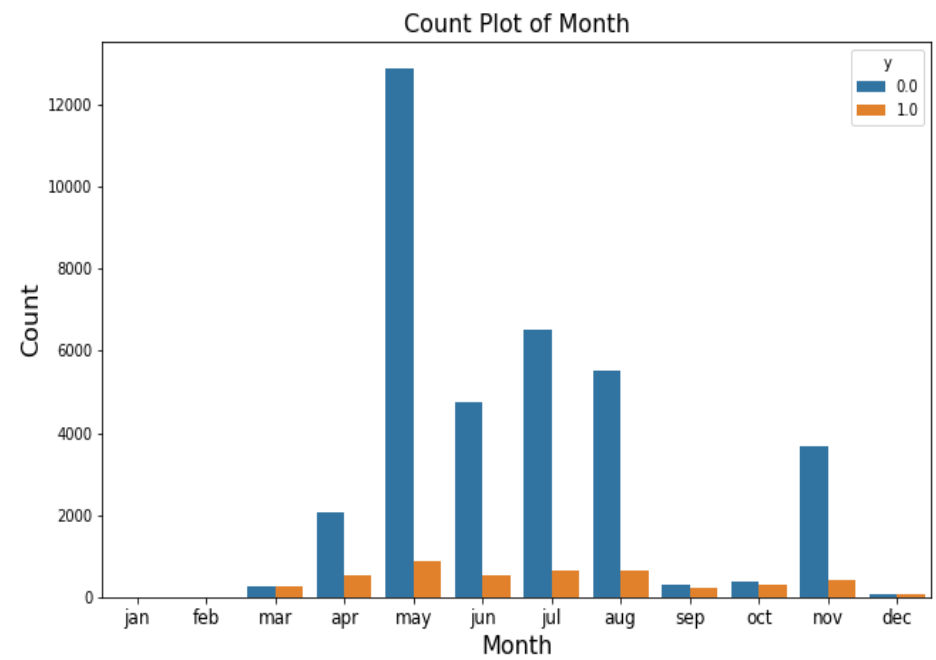
間隔天數(Pdays) - 特徵工程



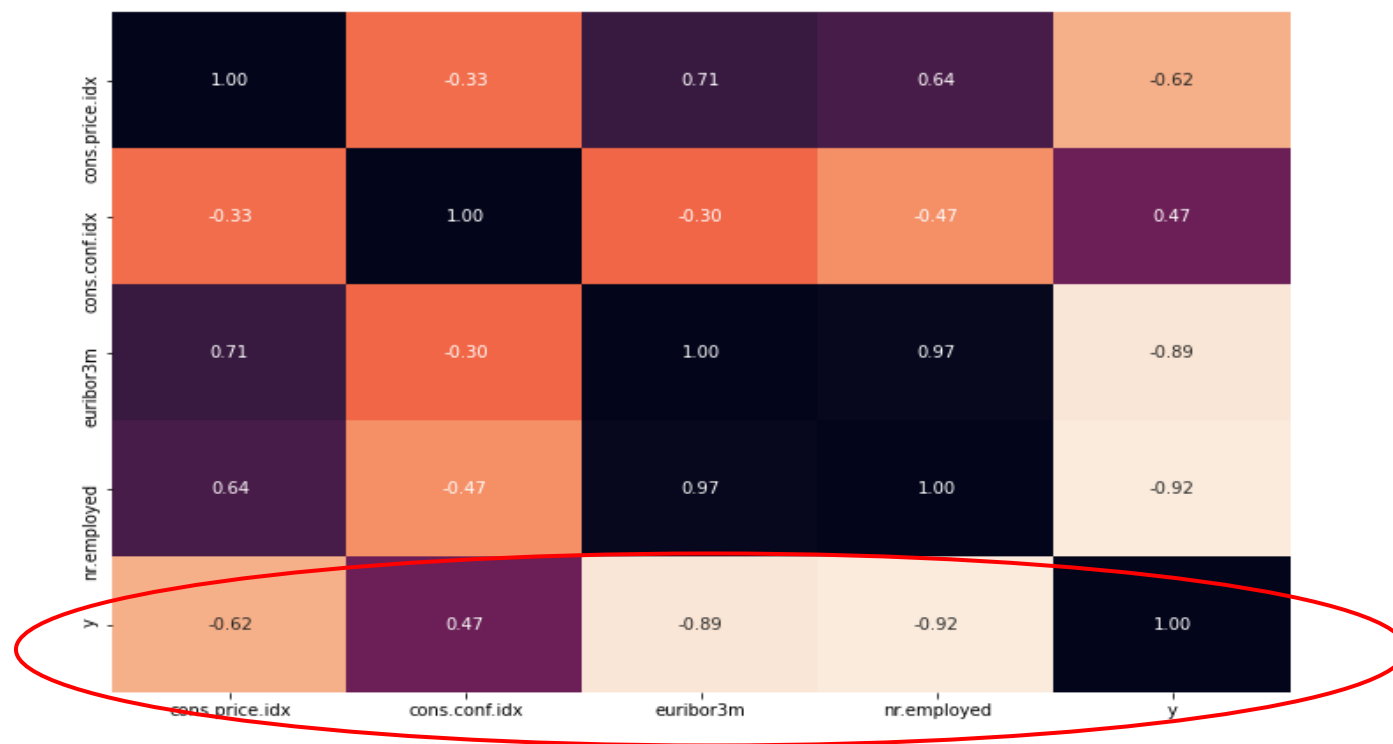
星期幾 (Day of Week)



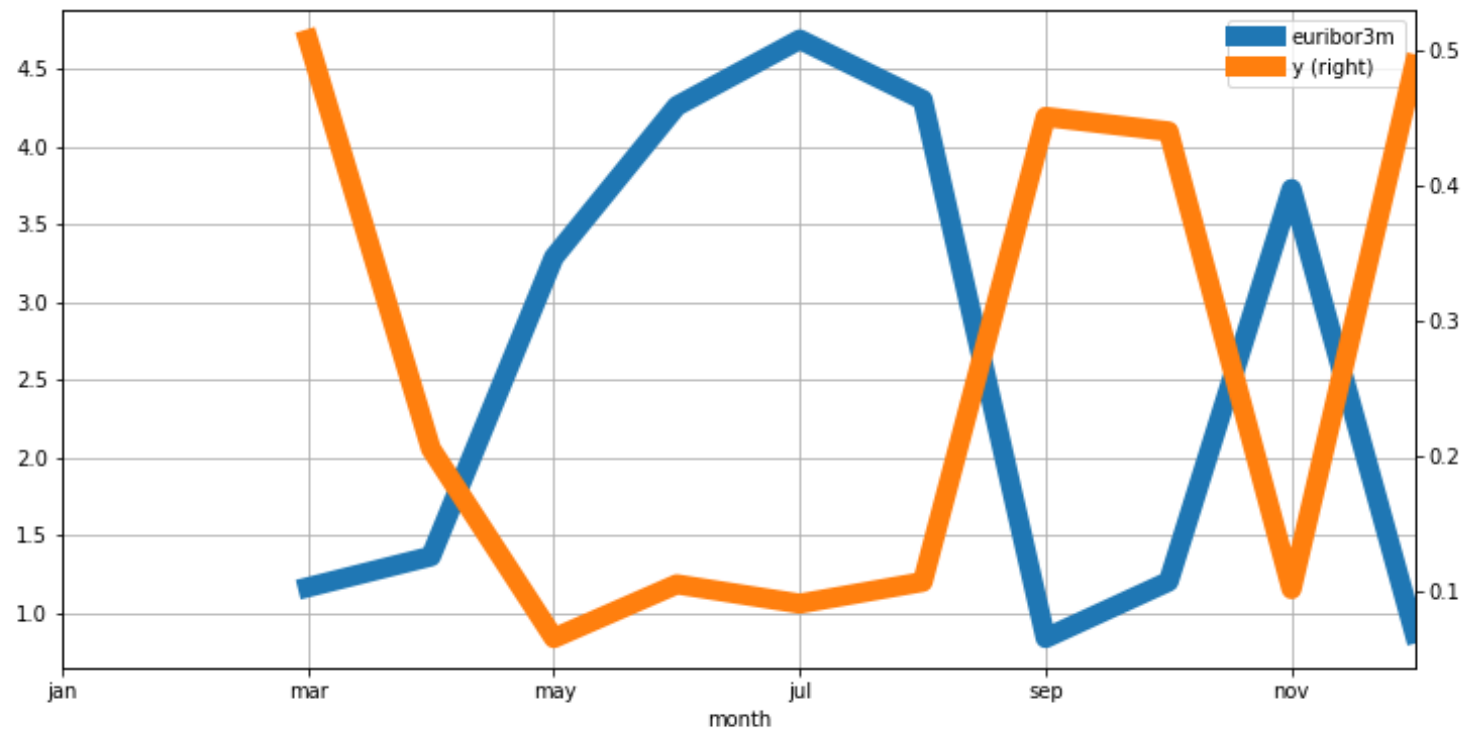
認購月份 (Month)



環境變數對目標的影響



同業拆放利率(euribor3m)



最終使用資料集 (N=51)

Train : Test = 0.75 : 0.25

campaign	: 0.0
pdays	: 1.0
cons.price.idx	: 93.994
cons.conf.idx	: -36.4
euribor3m	: 4.857
nr.employed	: 5191.0
y	: 0.0
job_blue-collar	: 0.0
job_entrepreneur	: 0.0
job_housemaid	: 1.0
job_management	: 0.0
job_retired	: 0.0
job_self-employed	: 0.0
job_services	: 0.0
job_student	: 0.0
job_technician	: 0.0
job_unemployed	: 0.0
marital_married	: 1.0
marital_single	: 0.0
education_basic.6y	: 0.0
education_basic.9y	: 0.0
education_high.school	: 0.0
education_illiterate	: 0.0
education_professional.course	: 0.0
education_university.degree	: 0.0

default_1	: 0.0
default_unknown	: 0.0
housing_1	: 0.0
housing_unknown	: 0.0
loan_1	: 0.0
loan_unknown	: 0.0
contact_telephone	: 1.0
month_aug	: 0.0
month_dec	: 0.0
month_jul	: 0.0
month_jun	: 0.0
month_mar	: 0.0
month_may	: 1.0
month_nov	: 0.0
month_oct	: 0.0
month_sep	: 0.0
day_of_week_mon	: 1.0
day_of_week_thu	: 0.0
day_of_week_tue	: 0.0
day_of_week_wed	: 0.0
previous_1	: 0.0
previous_2+	: 0.0
poutcome_nonexistent	: 1.0
poutcome_success	: 0.0
age_group_mid	: 1.0
age_group_old	: 0.0

評估方法

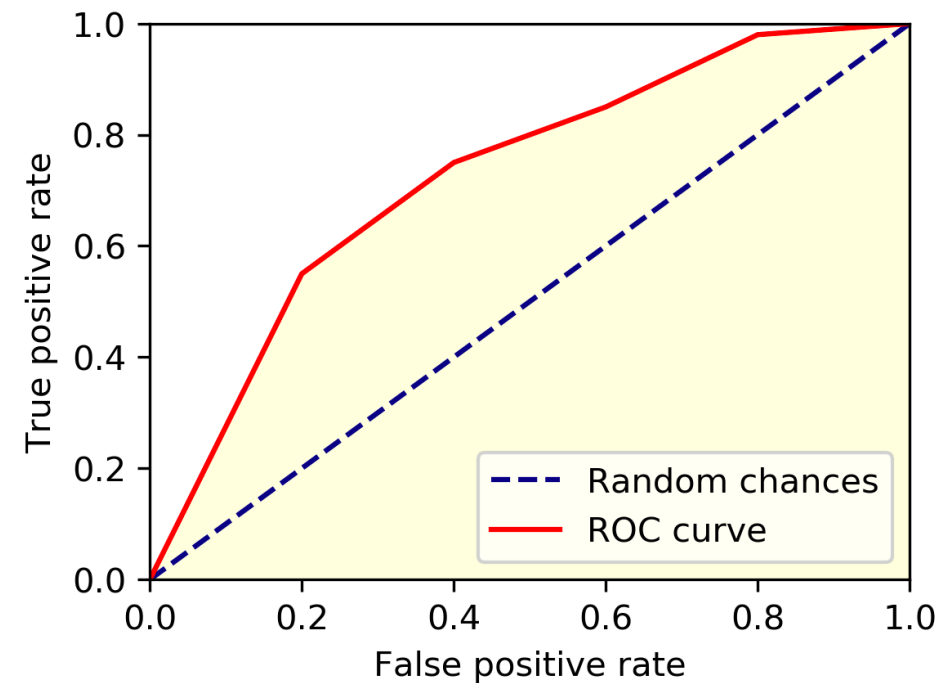
指標優化

► AUC (Area Under Curve)

	0	1
0	True Negative (TN)	False Positive (FP)
1	False Negative (FN)	True Positive (TP)

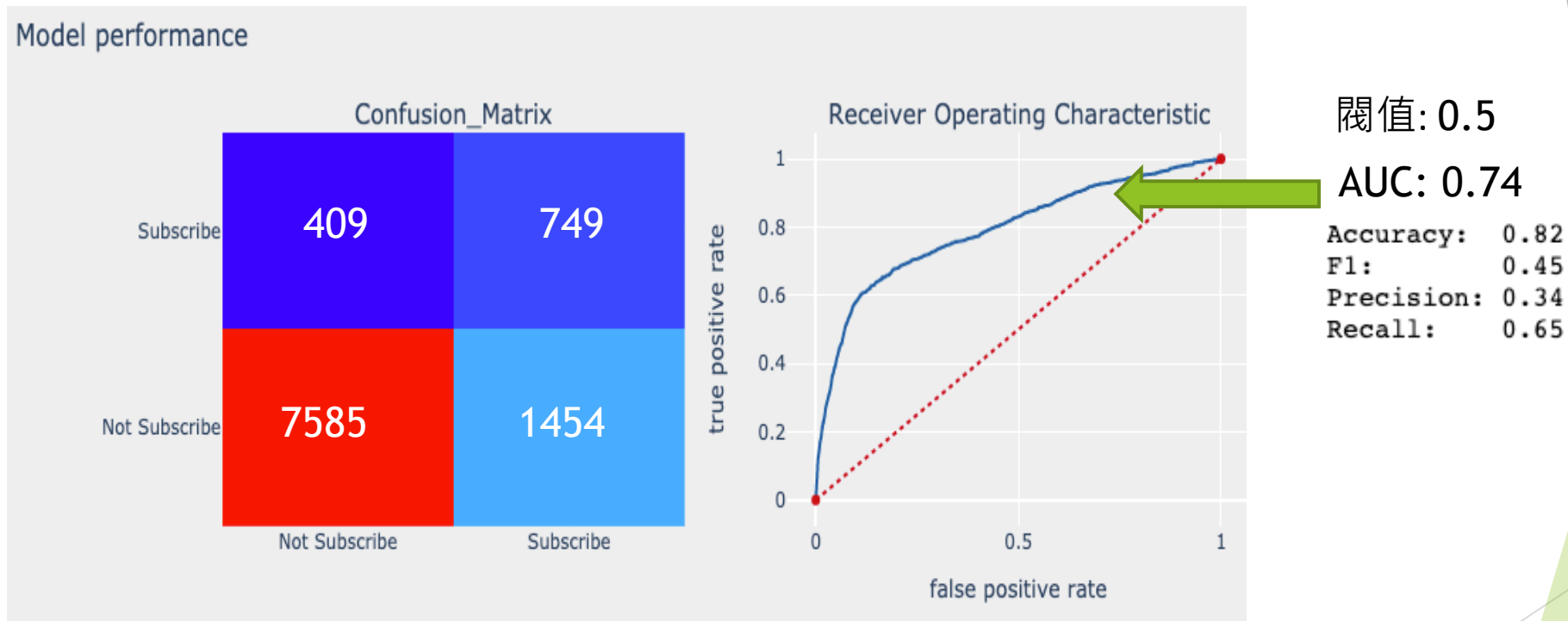
$$\text{TPR} = \text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{FPR} = \text{FP} / (\text{TN} + \text{FP})$$



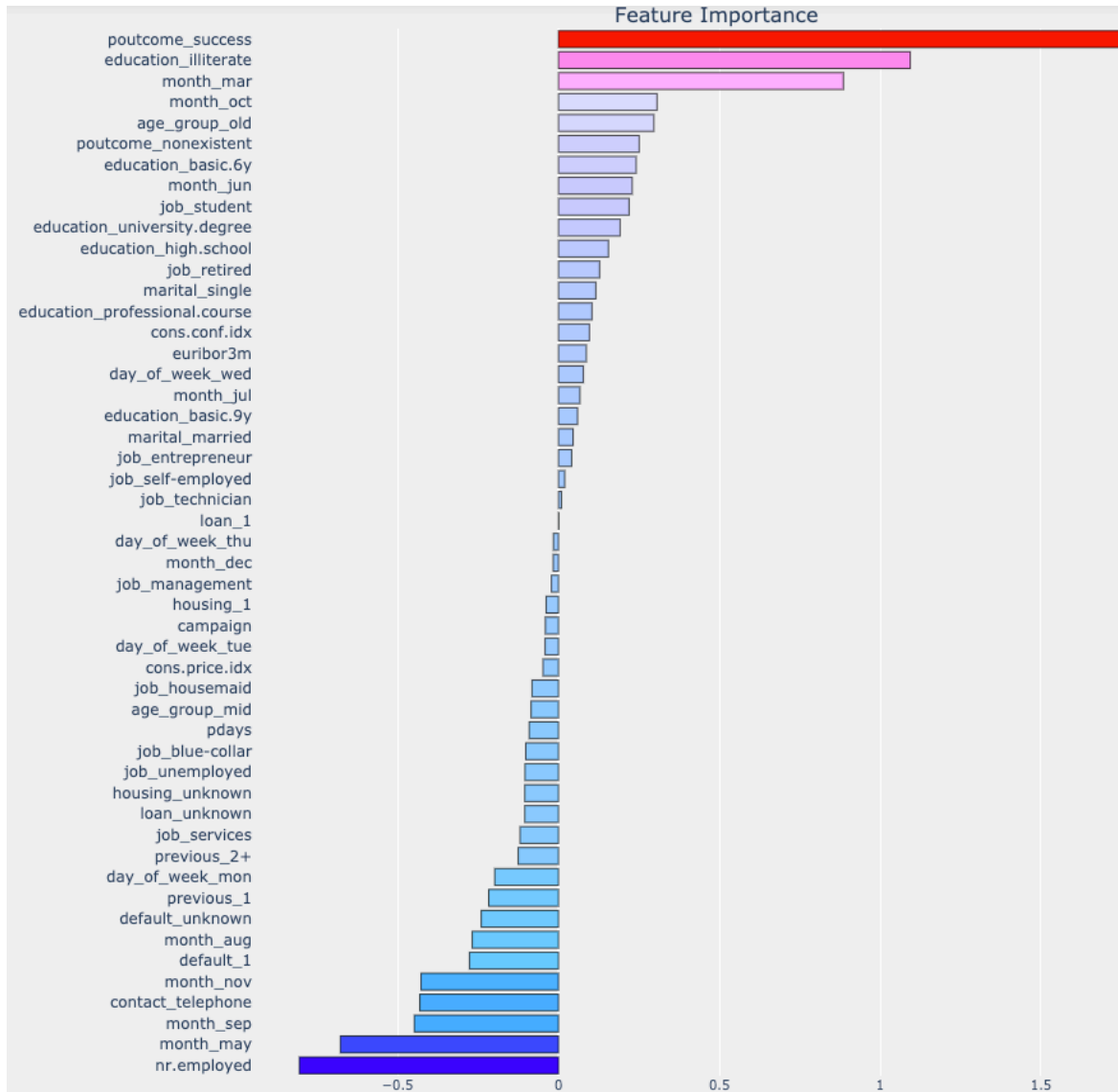
模型比較

邏輯斯回歸 (Logistic Regression)

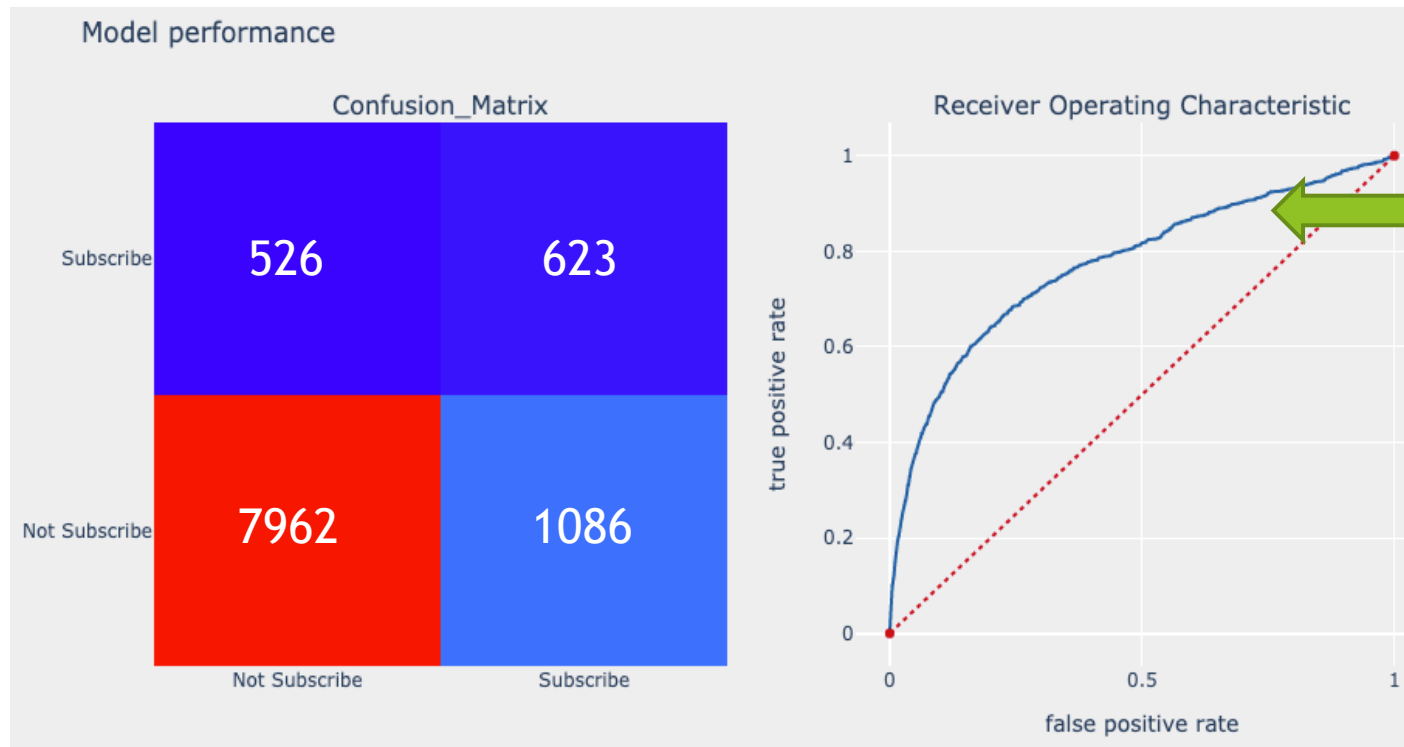


邏輯斯回歸 (Logistic Regression)

變數權重



隨機森林決策樹 (Random Forest)



閾值: 0.10

AUC: 0.77

Accuracy: 0.84

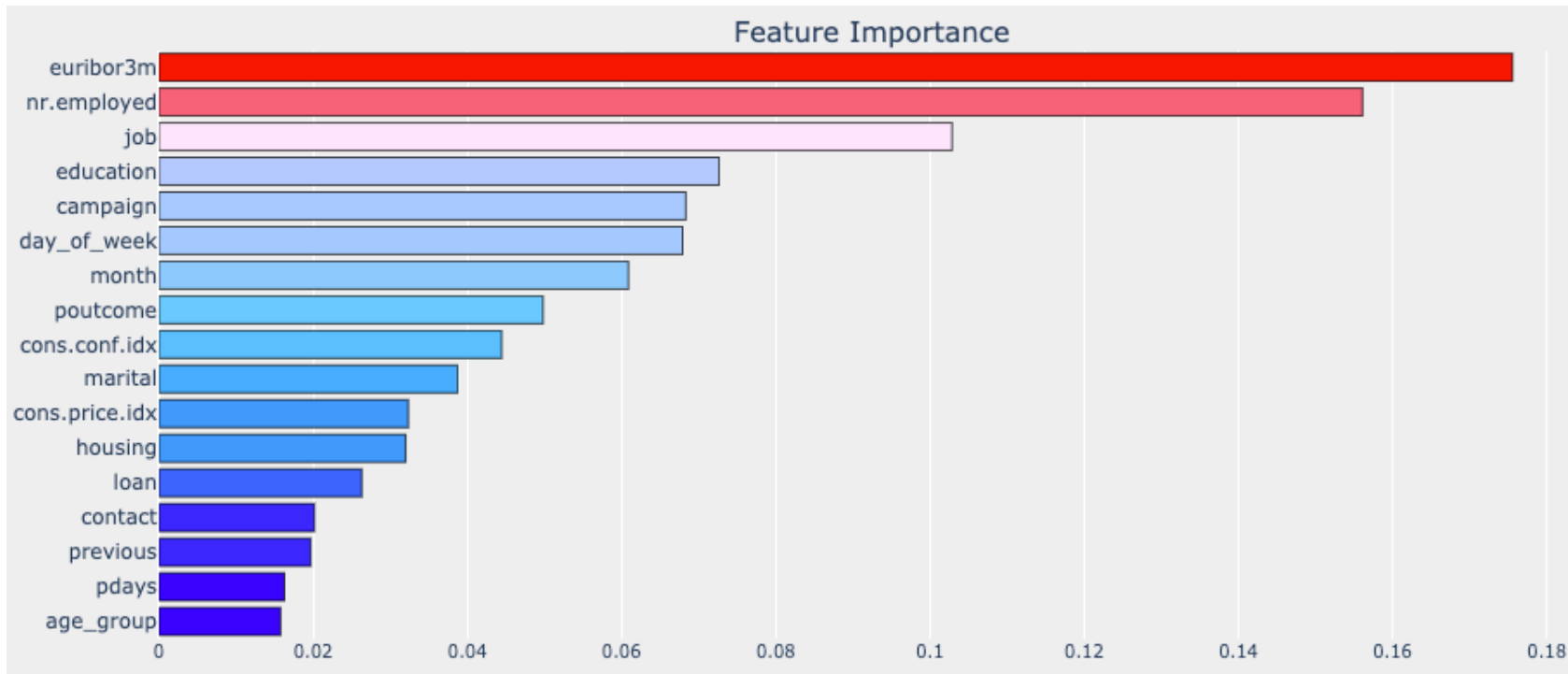
F1: 0.44

Precision: 0.36

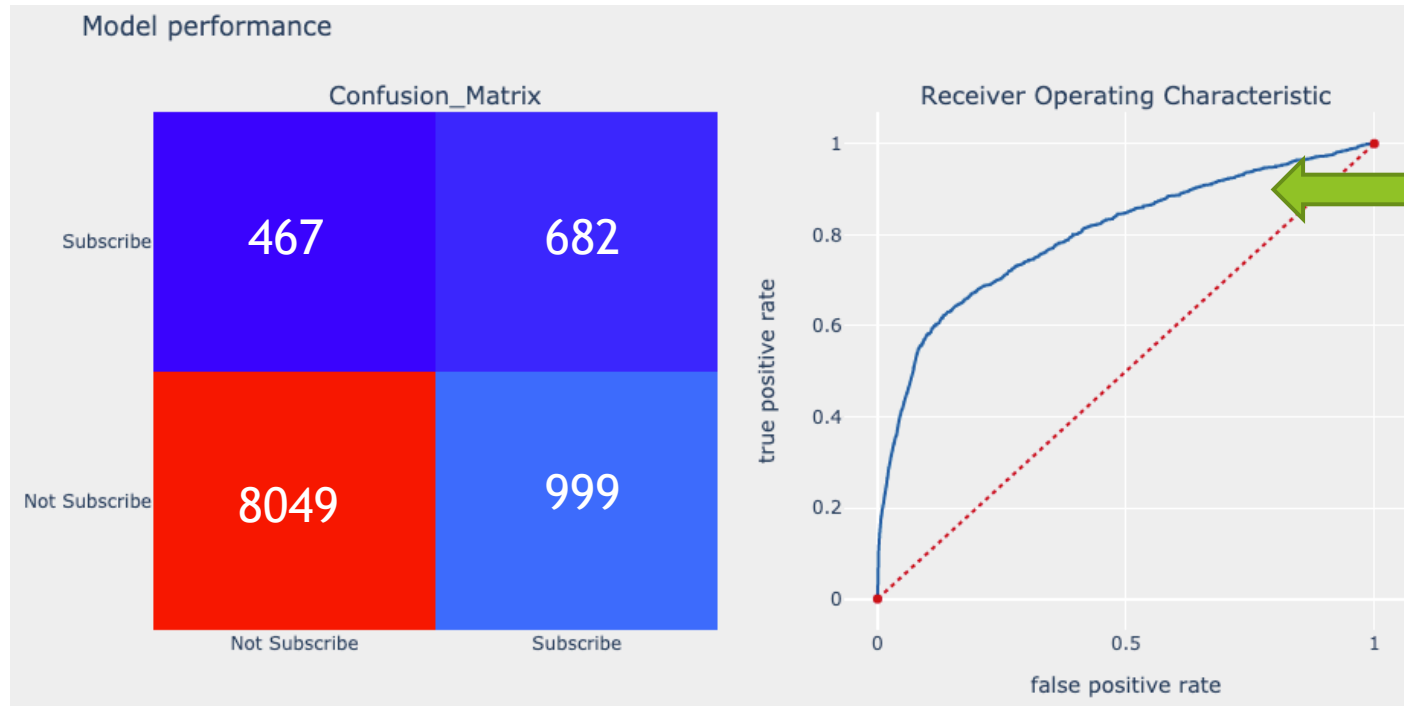
Recall: 0.54

隨機森林決策樹 (Random Forest)

變數重要性



梯度上升決策樹 (Gradient Boosting Decision Tree)



閾值: 0.13

AUC: 0.80

Accuracy: 0.86

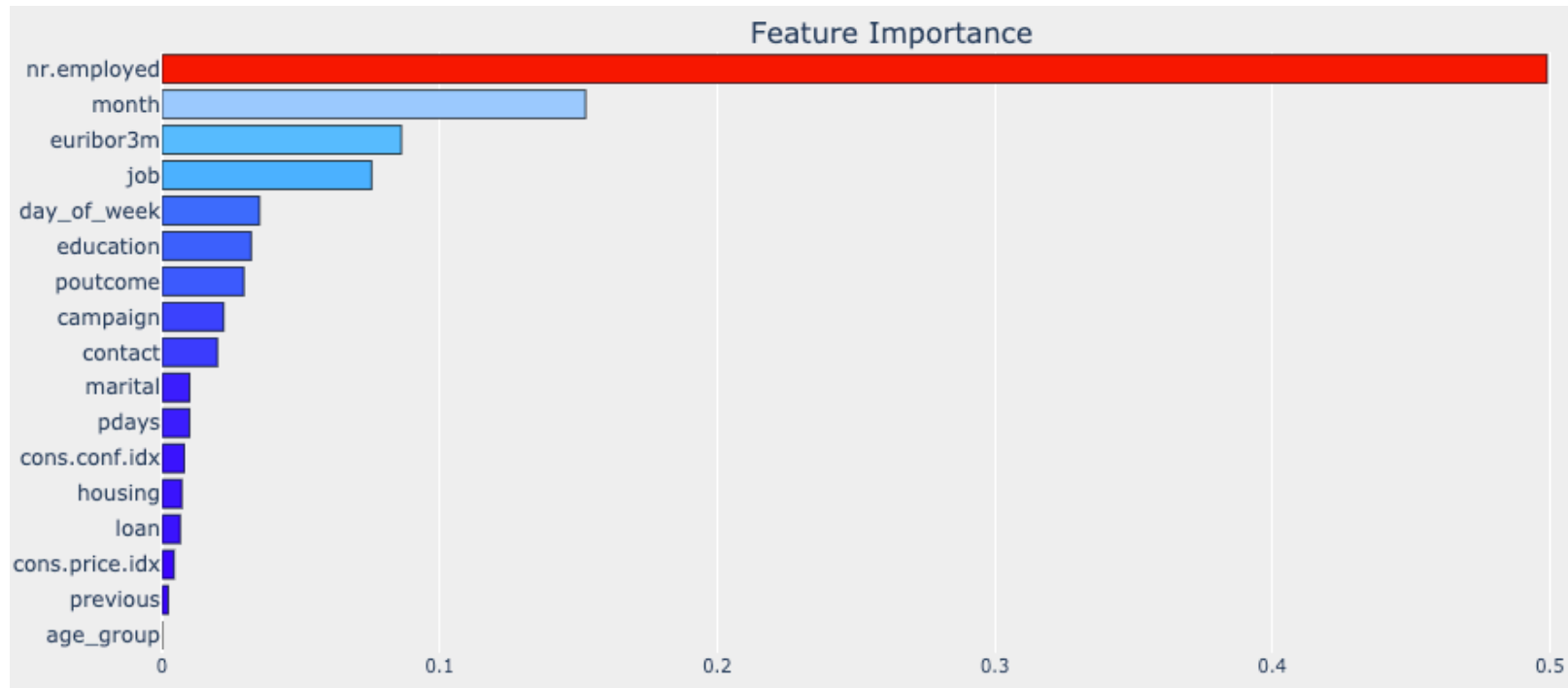
F1: 0.48

Precision: 0.41

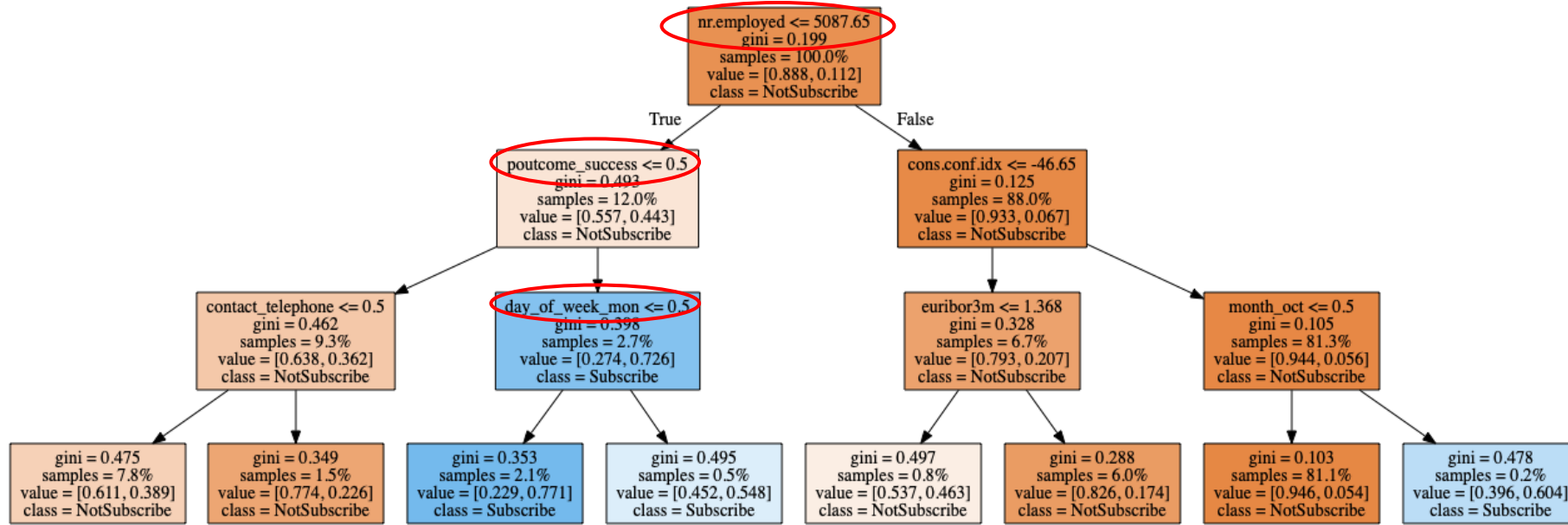
Recall: 0.59

梯度上升決策樹 (Gradient Boosting Decision Tree)

變數重要性



決策樹



客服準備撥打時的頁面

- ▶ 聯絡方式: (手機) 0800-036-599
- ▶ 客戶姓名: 國泰人
- ▶ 年齡: 20
- ▶ 工作: 學生
- ▶ 婚姻狀況: 單身
- ▶ 教育程度: 國中畢業
- ▶ 違約紀錄: 無
- ▶ 房子: 有
- ▶ 貸款: 無

- ▶ 此活動聯絡次數: 2
- ▶ 上次聯絡時間: 6
- ▶ 此活動前聯絡次數: 3
- ▶ 上次活動結果: 成功
- ▶ 持續時間: 361s

預測結果:

打!

07/04/2019 17:00

結論

1. 銀行行銷活動預測很明顯的會與整個市場環境有著很大的關聯，我們是否能有效透過時間序列來預測市場的走向成為非常重要的一環，如消費者物價或信心指數會如何變動，都是我們可以用來提高行銷活動成效的重要指標。
2. 因為遍佈各年齡層和不同的工作類別，我們需要去區分不同種類的用戶適合哪種行銷活動，可以透過質性或量性的調查來得知，建立強調個人化的行銷專案後也能透過A/B testing來驗證。
3. 在資料搜集方面可以更加全面，能加入更多的特徵進行加工和選擇，例如能加入撥打時間，地區和與客戶的交談內容等等使我們能更瞭解使用者行為。



國泰人壽

Cathay Life Insurance

Thank You!