請就所提供的資料進行以下分析：
1.分析持卡戶的特徵(性別、年齡、教育程度...等)對帳戶狀況(Attrition_Flag)之影響。
2.請以前述具顯著影響之特徵建立帳戶狀況(Attrition_Flag)之機器學習模型，並比較至少三種機器學習模型，分析後請以正確率(Accuracy)、精確率(Precision)、召回率(Recall)比較說明分析結果。

第一題

使用T檢定和卡方檢定，檢測「卡片類別、婚姻狀況、教育程度、所得、性別、與銀行合作月數、年齡、撫養人口、信用卡額度、持有卡片數、未使用卡片消費月數、總交易金額、信用卡可週轉額度、總交易筆數」中哪些對「帳戶狀況」有顯著影響。

從下表可以看到「所得、性別、信用卡額度、持有卡片數、未使卡片消費月數、總交易金額、信用卡可週轉額度、總交易筆數」對「帳戶狀況」有顯著影響，因為P值小於0.05，所以將這些變數設為自變數X。

卡方檢定				
變數	中文	卡方值	P值	有無顯著差異
Card_Category	卡片類別	2.234	0.525	無顯著差異
Marital_Status	婚姻狀況	6.056	0.109	無顯著差異
Education_Level	教育程度	12.511	0.051	無顯著差異
Income_Category	**所得**	**12.832**	**0.025**	**有顯著差異**
Gender	**性別**	**13.866**	**0.000**	**有顯著差異**
T檢定				
變數	中文	T值	P值	有無顯著差異
Months_on_book	與銀行合作月數	1.377	0.168	無顯著差異
Customer_Age	年齡	1.832	0.067	無顯著差異
Dependent_count	撫養人口	1.911	0.056	無顯著差異
Credit_Limit	**信用卡額度**	**-2.403**	**0.016**	**有顯著差異**
Total_Relationship_Count	**持有卡片數**	**-15.267**	**0.000**	**有顯著差異**
Months_Inactive_12_mon	**未使用卡片消費月數**	**15.521**	**0.000**	**有顯著差異**
Total_Trans_Amt	**總交易金額**	**-17.211**	**0.000**	**有顯著差異**
Total_Revolving_Bal	**信用卡可週轉額度**	**-27.435**	**0.000**	**有顯著差異**
Total_Trans_Ct	**總交易筆數**	**-40.251**	**0.000**	**有顯著差異**

第二題

使用決策樹、隨機森林、羅吉斯回歸三種機器學習來看各自的正確率、精確率、召回率

法一:自變數X取全部有顯著相關的變數，並利用決策樹、隨機森林、羅吉斯回歸三種機器學習方法

➔ 自變數X (全部有顯著相關的變數)

所得、性別、信用卡額度、持有卡片數、未使用卡片消費月數、總交易金額、信用卡可週轉額度、總交易筆數

➔ 依變數y

帳戶狀況

- 決策樹
  - 訓練
    - 正確率(Accuracy)=0.94
    - 精確率(Precision):已關閉=0.82、未關閉=0.96
    - 召回率(Recall):已關閉=0.78、未關閉=0.97
  - 測試
    - 正確率(Accuracy)=0.94
    - 精確率(Precision):已關閉=0.85、未關閉=0.96
    - 召回率(Recall):已關閉=0.79、未關閉=0.97

正確率接近1，且各項數值皆為正常數據，表此方法出來的混淆矩陣是有參考價值的。

決策樹分析的正確率、精確率、召回率在訓練跟測試出來的結果是相近的，且未關閉的數值皆高於已關閉。

決策樹分析的混淆矩陣

訓練

```
[[ 955  274]
 [ 204 6162]]
```

	precision	recall	f1-score	support
0	0.82	0.78	0.80	1229
1	0.96	0.97	0.96	6366
accuracy			0.94	7595
macro avg	0.89	0.87	0.88	7595
weighted avg	0.94	0.94	0.94	7595

測試

```
[[ 314   84]
 [  56 2078]]
```

	precision	recall	f1-score	support
0	0.85	0.79	0.82	398
1	0.96	0.97	0.97	2134
accuracy			0.94	2532
macro avg	0.90	0.88	0.89	2532
weighted avg	0.94	0.94	0.94	2532

註:帳戶狀況:已關閉=0,未關閉=1

- 隨機森林
  - 訓練
    - 正確率(Accuracy)=1.00
    - 精確率(Precision):已關閉=1.00、未關閉=1.00
    - 召回率(Recall):已關閉=1.00、未關閉=1.00
  - 測試
    - 正確率(Accuracy)=0.95
    - 精確率(Precision):已關閉=0.89、未關閉=0.96
    - 召回率(Recall):已關閉=0.79、未關閉=0.98

正確率接近1，但在訓練時所有項目的數值都等於1，表此方法出來的混淆矩陣可能是有問題的，所以即便正確率接近1，此方法相較前個方法較不適合拿來參考。

隨機森林的混淆矩陣

訓練

```
[[1229    0]
 [   0 6366]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      1229
           1       1.00      1.00      1.00      6366

    accuracy                           1.00      7595
   macro avg       1.00      1.00      1.00      7595
weighted avg       1.00      1.00      1.00      7595
```

測試

```
[[ 314   84]
 [  37 2097]]
              precision    recall  f1-score   support

           0       0.89      0.79      0.84       398
           1       0.96      0.98      0.97      2134

    accuracy                           0.95      2532
   macro avg       0.93      0.89      0.91      2532
weighted avg       0.95      0.95      0.95      2532
```

註:帳戶狀況:已關閉=0,未關閉=1

- 羅吉斯回歸
  - 訓練
    - 正確率(Accuracy)=0.87
    - 精確率(Precision):已關閉=0.70、未關閉=0.88
    - 召回率(Recall):已關閉=0.34、未關閉=0.97
  - 測試
    - 正確率(Accuracy)=0.87
    - 精確率(Precision):已關閉=0.67、未關閉=0.89
    - 召回率(Recall):已關閉=0.34、未關閉=0.97

羅吉斯回歸的正確率、精確率、召回率在訓練跟測試出來的結果是相近的，但已關閉的值小於0.5可靠性低，所以即便正確率接近1，此方法相較前兩個方法較不適合拿來參考。

羅吉斯迴歸的混淆矩陣

訓練

```
[[ 413  816]
 [ 180 6186]]
              precision    recall  f1-score   support

           0       0.70      0.34      0.45      1229
           1       0.88      0.97      0.93      6366

    accuracy                           0.87      7595
   macro avg       0.79      0.65      0.69      7595
weighted avg       0.85      0.87      0.85      7595
```

測試

```
[[ 135  263]
 [  67 2067]]
              precision    recall  f1-score   support

           0       0.67      0.34      0.45       398
           1       0.89      0.97      0.93      2134

    accuracy                           0.87      2532
   macro avg       0.78      0.65      0.69      2532
weighted avg       0.85      0.87      0.85      2532
```

註:帳戶狀況:已關閉=0,未關閉=1

以取全部有顯著相關的變數來看，最值得拿來使用的是決策樹，因為正確率、精確率、召回率的數值合理性優於另外兩種，比較有利用價值。

法二:自變數X取類別和屬量變數相對更有顯著相關的變數前兩名，並利用決策樹、隨機森林、羅吉斯回歸三種機器學習方法

➔ 自變數X (類別和屬量變數相對更有顯著相關的變數前兩名)

所得、性別、信用卡可週轉額度、總交易筆數

➔ 依變數y

帳戶狀況

- 決策樹
  - 訓練
    - 正確率(Accuracy)=0.90
    - 精確率(Precision):已關閉=0.74、未關閉=0.92
    - 召回率(Recall):已關閉=0.56、未關閉=0.96
  - 測試
    - 正確率(Accuracy)=0.90
    - 精確率(Precision):已關閉=0.75、未關閉=0.93
    - 召回率(Recall):已關閉=0.59、未關閉=0.96

正確率接近1，但已關閉的召回率不論在測試還是訓練都不太好，所以此方法出來的混淆矩陣的參考價值有待觀察。

決策樹分析的混淆矩陣

訓練

```
[[ 693  536]
 [ 243 6123]]
              precision    recall  f1-score   support

           0       0.74      0.56      0.64      1229
           1       0.92      0.96      0.94      6366

    accuracy                           0.90      7595
   macro avg       0.83      0.76      0.79      7595
weighted avg       0.89      0.90      0.89      7595
```

測試

```
[[ 233  165]
 [  79 2055]]
              precision    recall  f1-score   support

           0       0.75      0.59      0.66       398
           1       0.93      0.96      0.94      2134

    accuracy                           0.90      2532
   macro avg       0.84      0.77      0.80      2532
weighted avg       0.90      0.90      0.90      2532
```

註:帳戶狀況:已關閉=0,未關閉=1

- 隨機森林
  - 訓練
    - 正確率(Accuracy)=0.96
    - 精確率(Precision):已關閉=0.89、未關閉=0.97
    - 召回率(Recall):已關閉=0.86、未關閉=0.98
  - 測試
    - 正確率(Accuracy)=0.88
    - 精確率(Precision):已關閉=0.64、未關閉=0.92
    - 召回率(Recall):已關閉=0.56、未關閉=0.94

正確率接近1，但已關閉的精確率和召回率在測試跟訓練都不太好，所以此方法出來的混淆矩陣的參考價值有待觀察。

隨機森林的混淆矩陣	
訓練	測試

[[1063 166] [125 6241]] ========== precision recall f1-score support 0 0.89 0.86 0.88 1229 1 0.97 0.98 0.98 6366 accuracy 0.96 7595 macro avg 0.93 0.92 0.93 7595 weighted avg 0.96 0.96 0.96 7595	[[224 174] [125 2009]] ========== precision recall f1-score support 0 0.64 0.56 0.60 398 1 0.92 0.94 0.93 2134 accuracy 0.88 2532 macro avg 0.78 0.75 0.77 2532 weighted avg 0.88 0.88 0.88 2532

註:帳戶狀況:已關閉=0,未關閉=1

- 羅吉斯回歸
  - 訓練
    - 正確率(Accuracy)=0.87
    - 精確率(Precision):已關閉=0.69、未關閉=0.89
    - 召回率(Recall):已關閉=0.37、未關閉=0.97
  - 測試
    - 正確率(Accuracy)=0.87
    - 精確率(Precision):已關閉=0.66、未關閉=0.89
    - 召回率(Recall):已關閉=0.39、未關閉=0.96

羅吉斯回歸的正確率、精確率、召回率在訓練跟測試出來的結果是相近的，但未關閉的值小於0.5可靠性低，所以即便正確率接近1，此方法相較前兩個方法較不適合拿來參考。

羅吉斯迴歸的混淆矩陣	
訓練	測試
[[450 779] [199 6167]] ========== precision recall f1-score support 0 0.69 0.37 0.48 1229 1 0.89 0.97 0.93 6366 accuracy 0.87 7595 macro avg 0.79 0.67 0.70 7595 weighted avg 0.86 0.87 0.85 7595	[[155 243] [80 2054]] ========== precision recall f1-score support 0 0.66 0.39 0.49 398 1 0.89 0.96 0.93 2134 accuracy 0.87 2532 macro avg 0.78 0.68 0.71 2532 weighted avg 0.86 0.87 0.86 2532

註:帳戶狀況:已關閉=0,未關閉=1

結論

綜合上述三種機器學習方法以及自變數的不同選擇，最值得拿來使用的是利用決策樹以及取全部有顯著相關的變數作為自變數的方法，因為他的正確率、精確率、召回率的數值合理性優於其他方法，比較有利用價值。