



# 機器學習-第四組

- *Second* -

組員:

巨資碩一	09370008	王怡人
巨資碩一	09370009	張文綺
經濟碩一	09351010	陳博勁
經濟碩一	09351012	張耀仁



# CONTENTS

## ■ PART 1

問題的定義&說明

## ■ PART 2

資料來源&前處理步驟

## ■ PART 3

資料EDA&視覺化解讀

## ■ PART 4

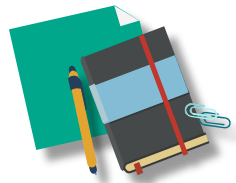
初步結果說明

## ■ PART 5

會議記錄

1

## 問題的定義&說明

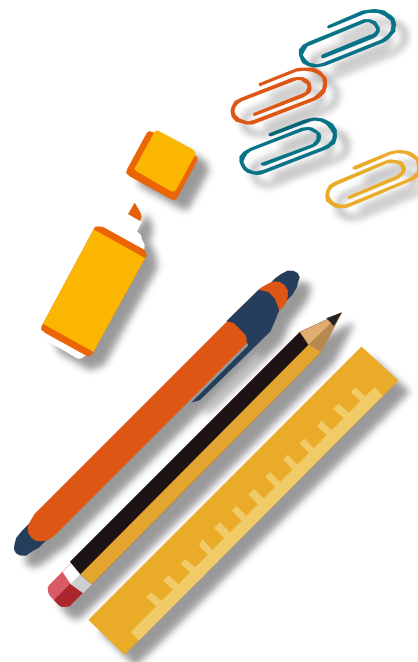


# 電子業就是夯！這家公司值得投資嗎？

## 財報總是看不懂？想投資不知道如何下手嗎？

本專題以營業毛利率去區分這家公司是否賺錢（值得投資），然後找出比較值得參考的指標，讓新手投資者可以藉由這些指標簡單的做判斷，作為入門的幫助。

- 選取的指標包含：經營能力相關指標 & 電子業特殊指標
- 資料：選擇疫情前(2019)年的財報
- 資料來源：T E J

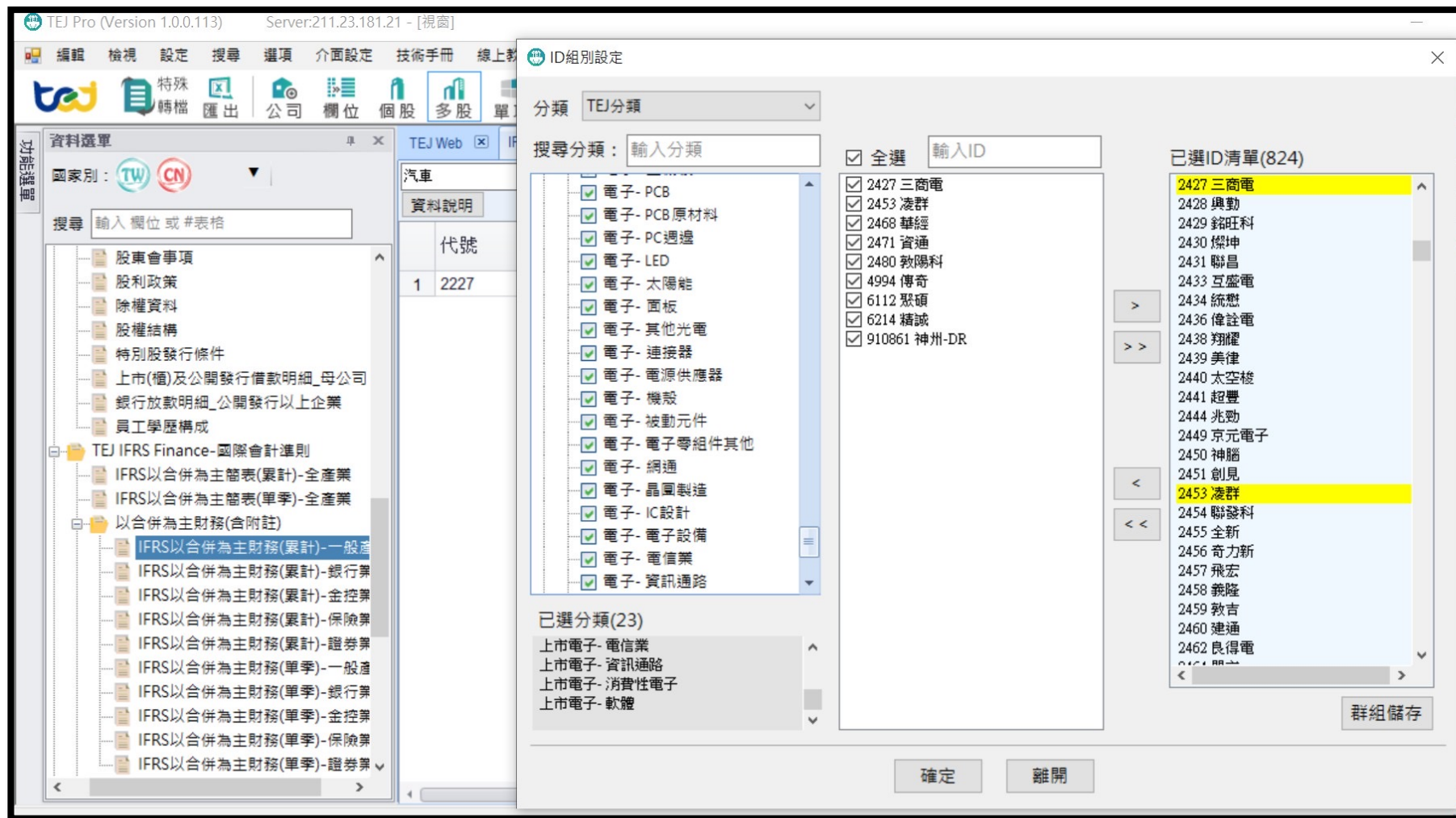


2

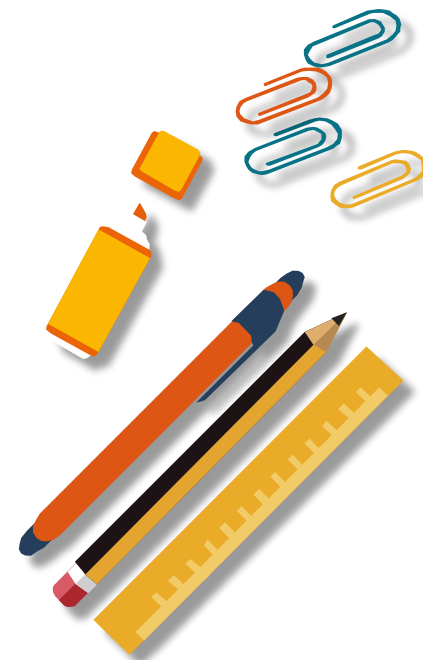
## 資料來源&前處理步驟



# TEJ 資料篩選 & 輸出



# 選擇電子產業



X1 . X2.....

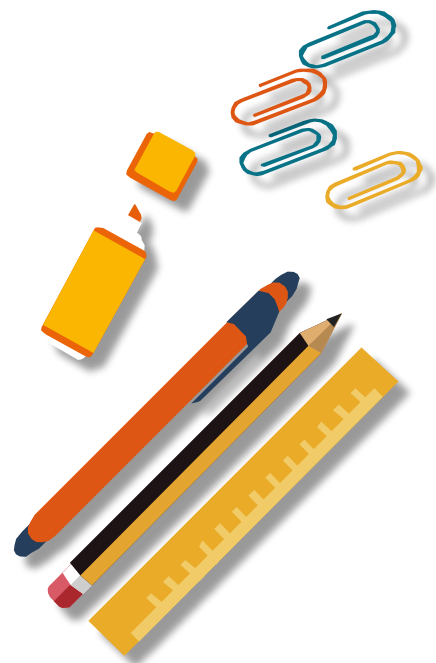
Y

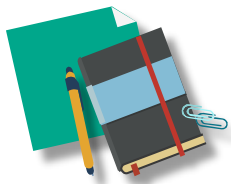
代號	名稱	年/月	營業毛 利率	研究發展 費用率	負債 比率	淨營業週期 (日)	平均售貨 天數	平均收 帳天數	總資產週 轉次數	固定資產 週轉次數	存貨週轉率 (次)	速動比 率	應收帳款 週轉次數	營業費 用率	流動比 率	總資產週轉 率	應收帳款週 轉率	應付帳款 週轉率
1435	中福	2019-12-01	94.03	0.00	15.73	1643.97	1638.75	21.34	0.02	0.23	0.22	11701.98	17.10	173.99	11882.10	1.974891	1.314178	2.604578
1471	首利	2019-12-01	-6.32	1.79	48.75	89.70	28.37	122.24	0.63	2.04	12.86	77.58	2.99	27.69	97.48	73.481958	22.356747	13.403432
1537	廣陸	2019-12-01	23.38	0.51	25.40	130.39	98.06	54.95	1.41	7.54	3.72	273.01	6.64	6.83	422.62	145.792741	14.248066	4.309813

企業能力	代表指標
電子產業特殊比率	研發費用率
償債能力	負債比率.流動比率.速動比率
經營能力	存貨週轉率.應付帳款週轉率.應收帳款週轉率.固定資產週轉率. 營業週期.營業費用率.總資產週轉率.平均收帳天數
獲利能力	營業毛利率

相關公式:

- 營業週期 → 平均銷貨日數+平均收現日數
- 營業費用率 → 營業費用/銷貨淨額

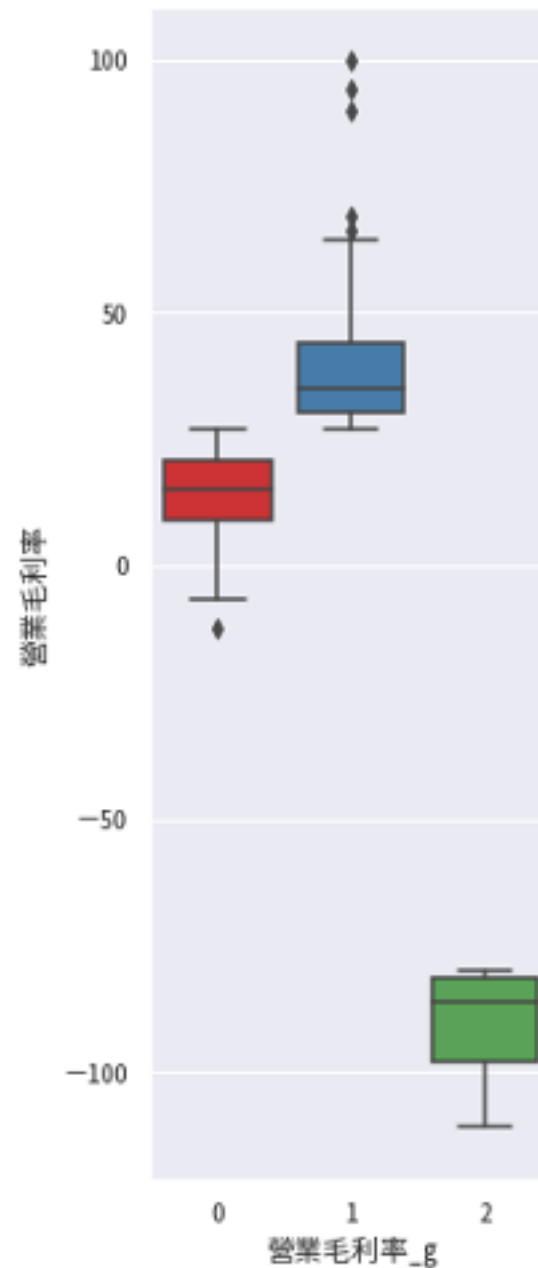




## Y 換成 “營業毛利率”

利用 K-means 把營業毛利率分成三組後將分組當成 Y，  
把公司分組看看各公司在哪一組。

營業毛利率_g		0	1	2
營業毛利率	count	288.000000	119.000000	5.000000
	mean	14.458819	39.168235	-91.210000
	std	7.808814	12.775372	12.940029
	min	-12.110000	26.980000	-110.640000
	25%	8.855000	30.405000	-97.880000
	50%	15.320000	35.240000	-86.150000
	75%	20.907500	44.020000	-81.330000
	max	26.810000	99.830000	-80.050000





# 3

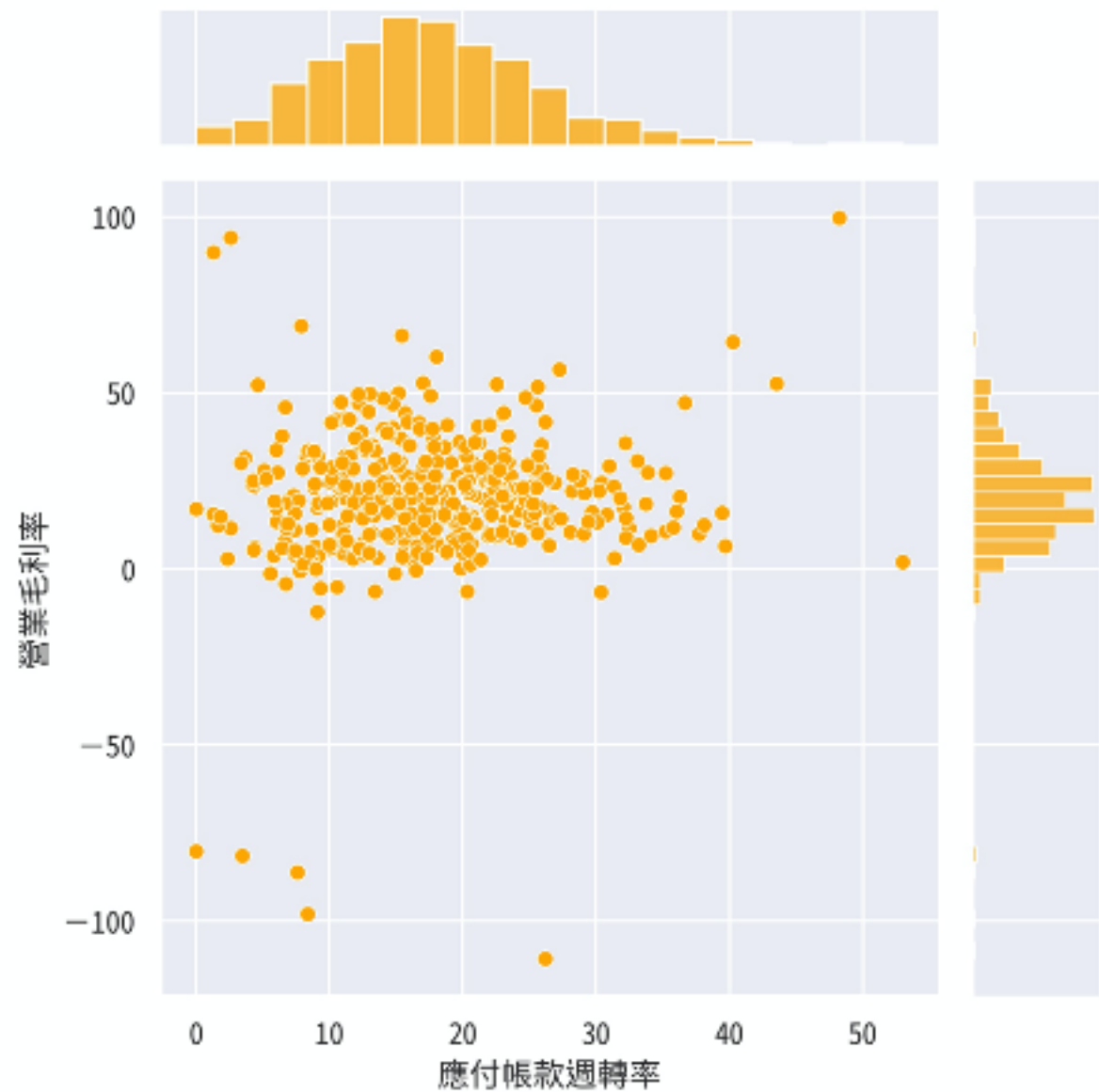
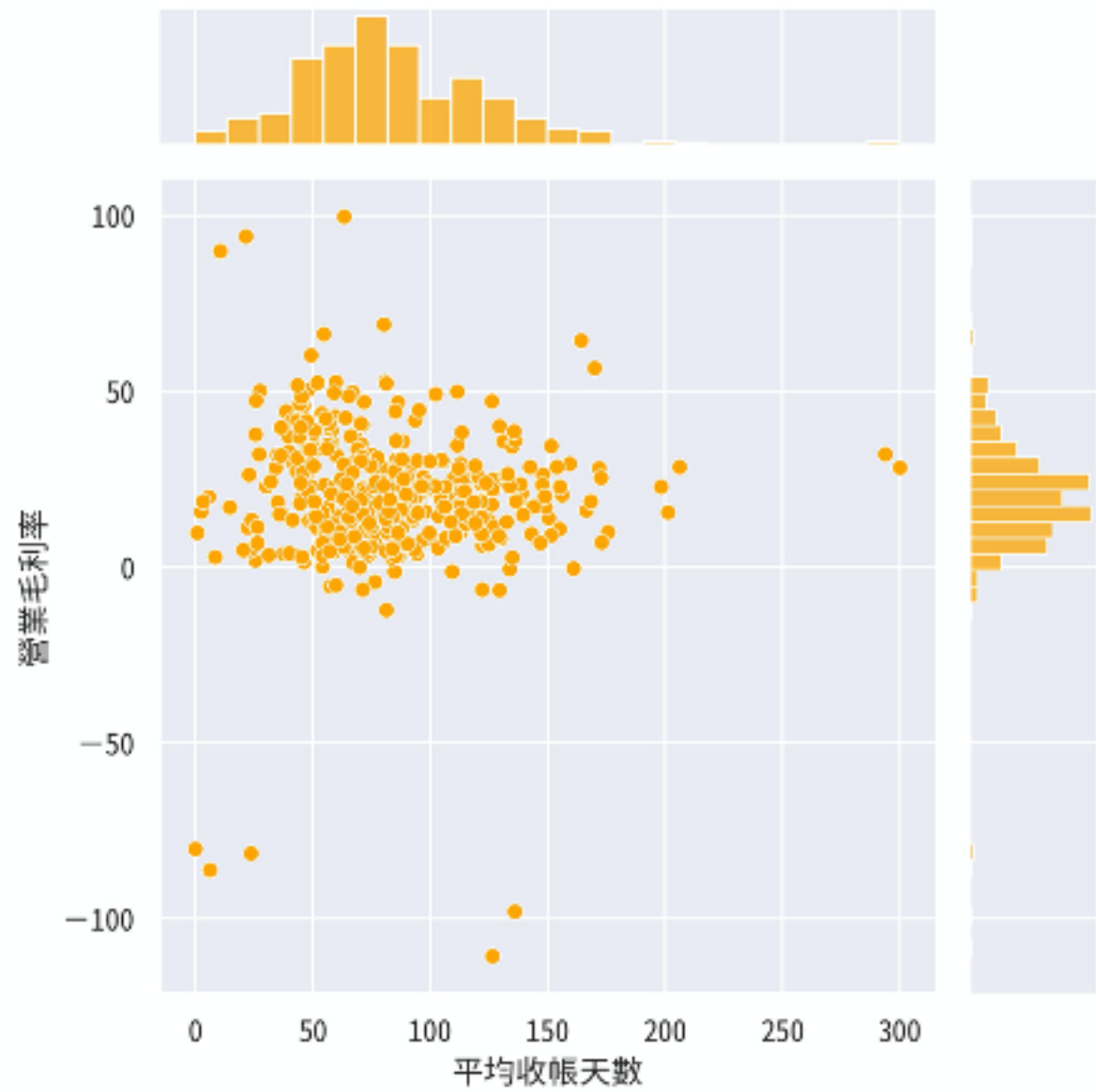
## 資料EDA & 視覺化解讀

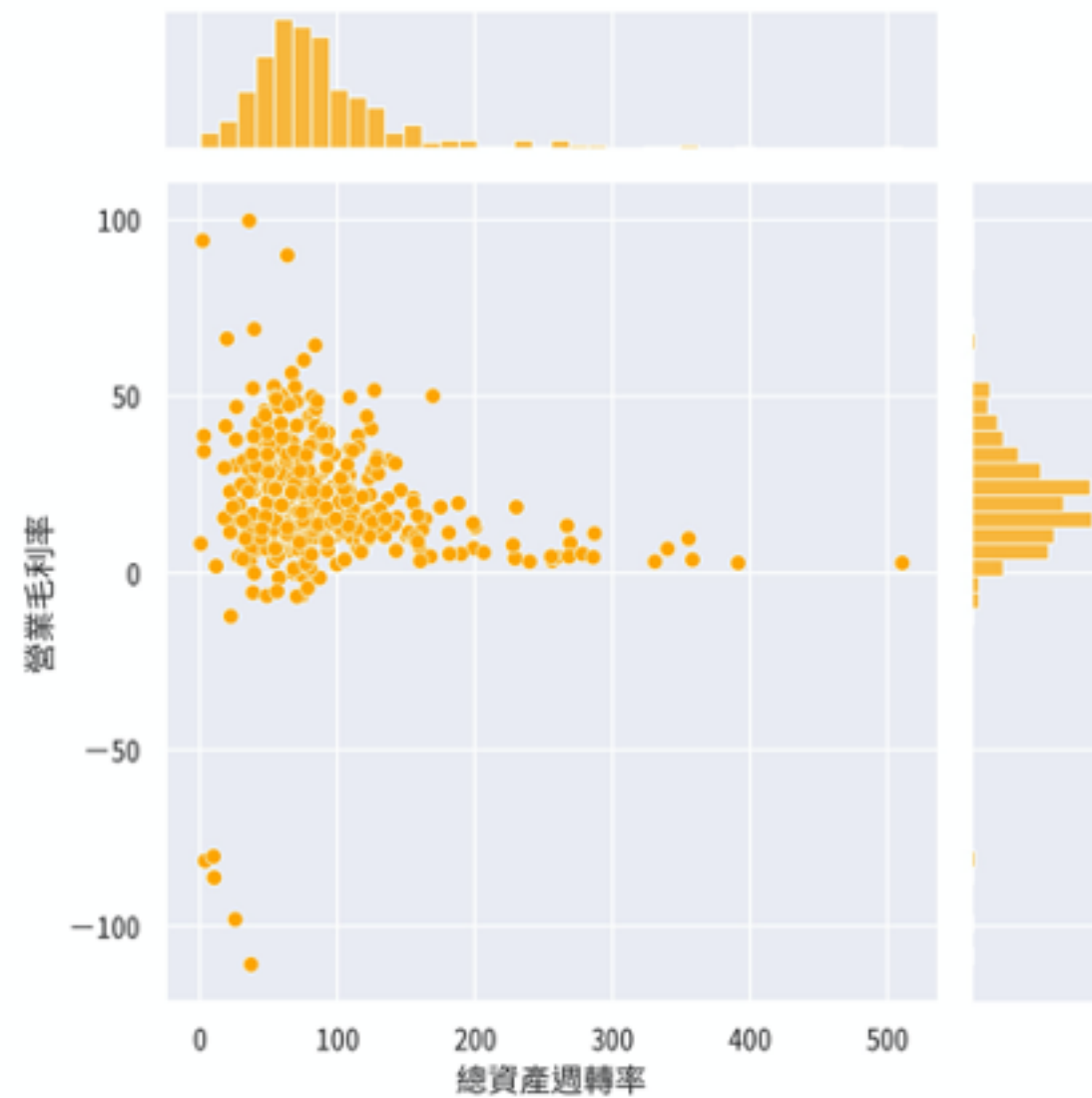
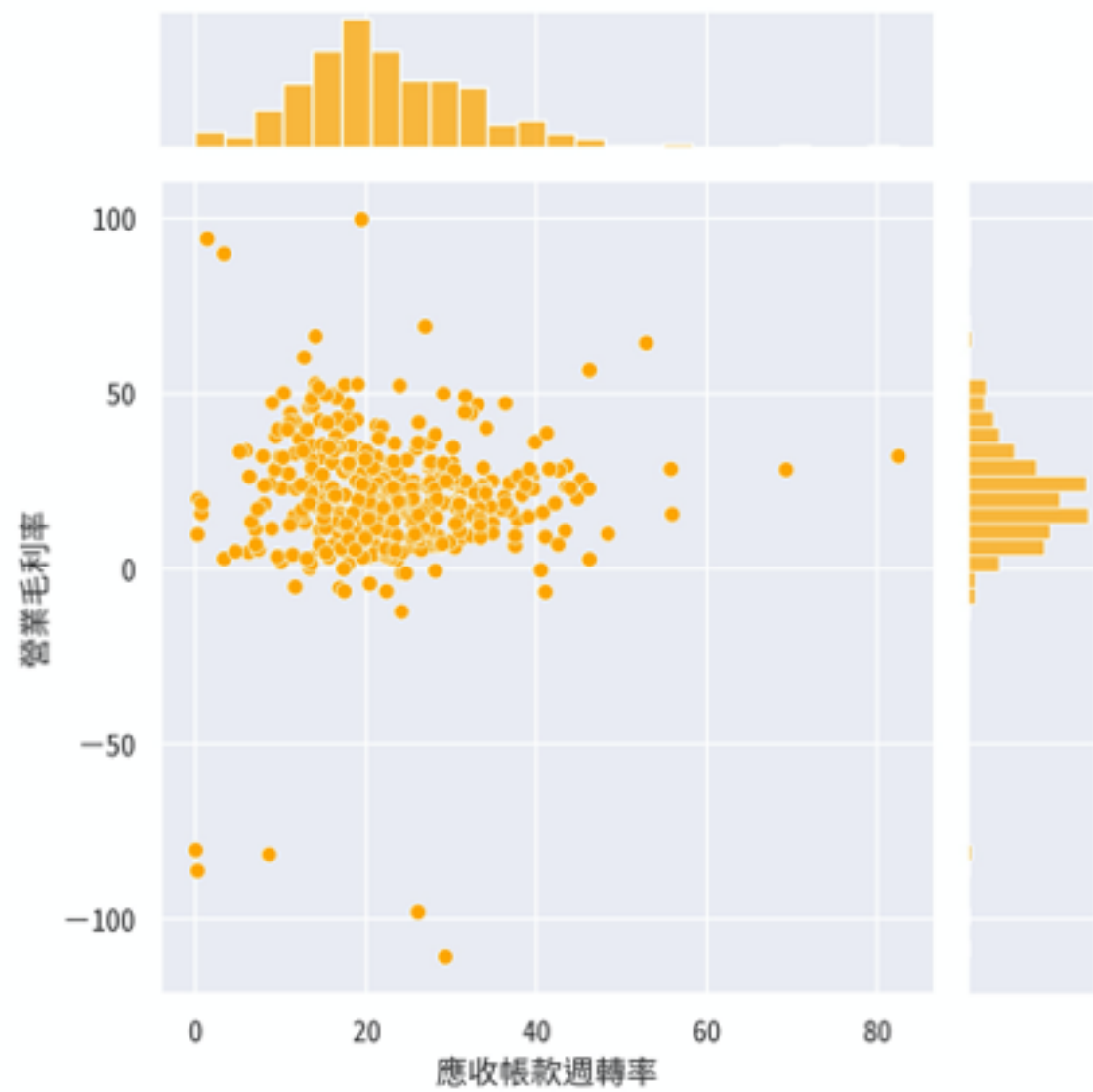
- 營業毛利率和各指標間的關係圖
- 各變數間相關係數圖

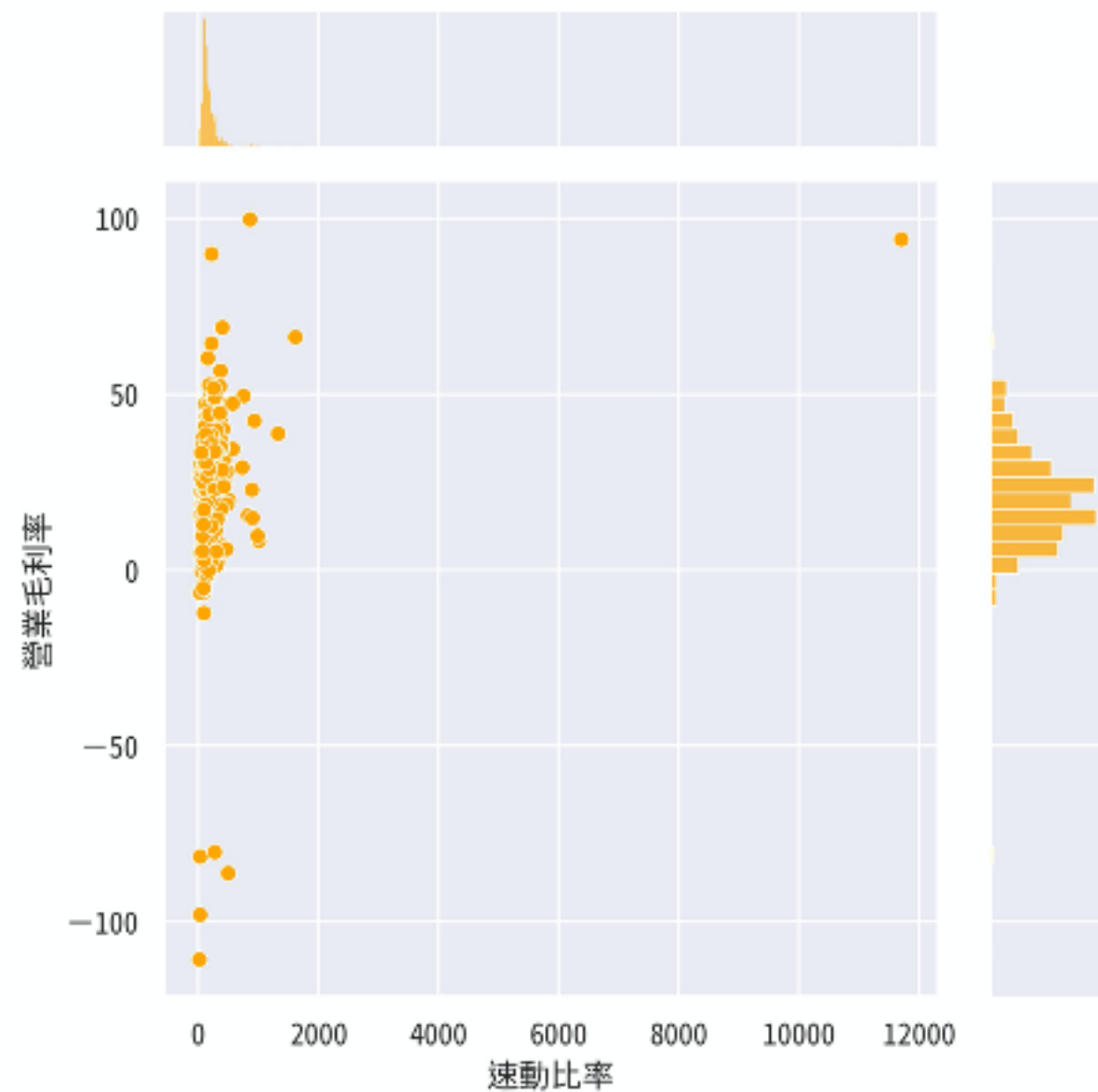
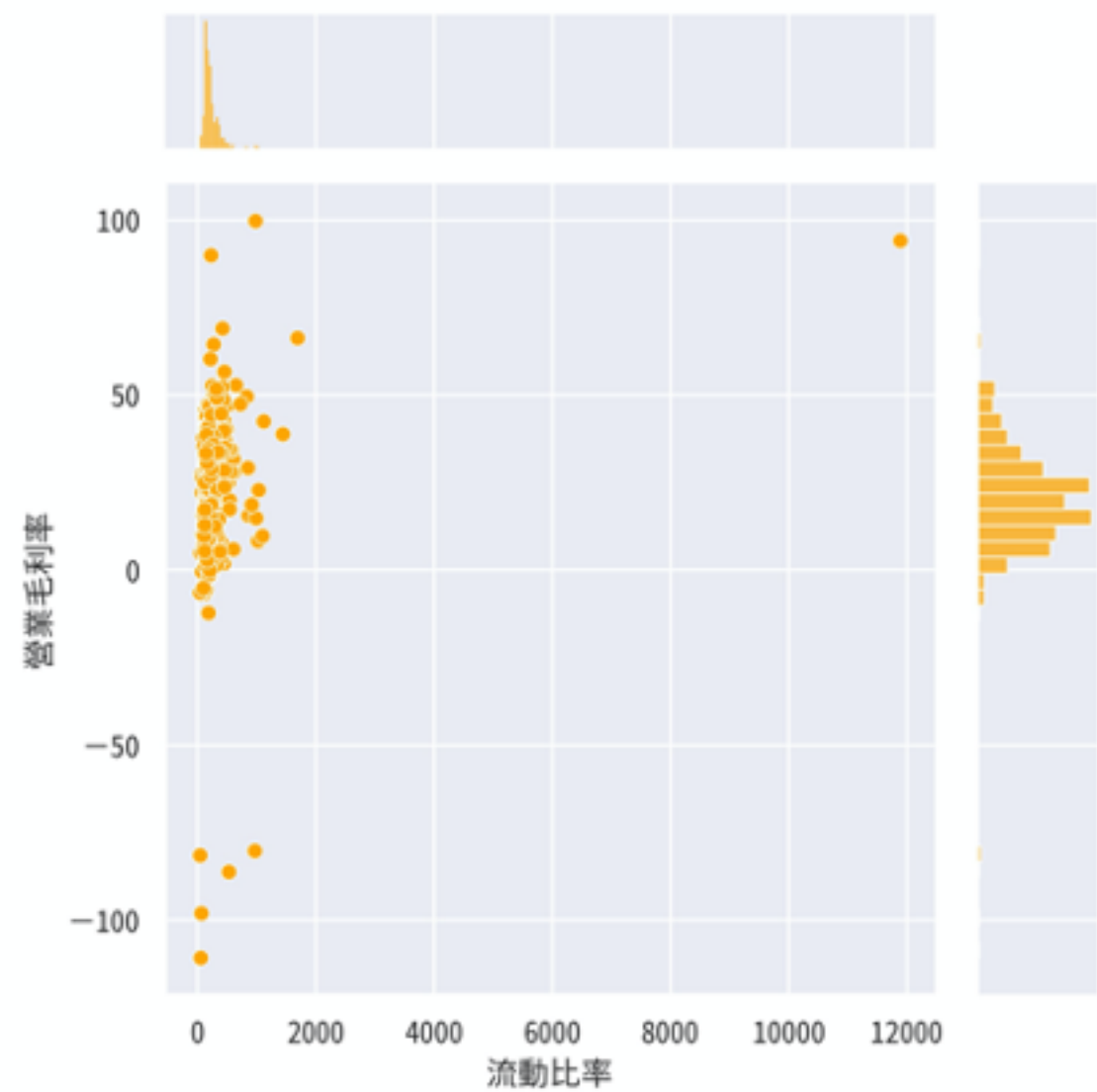


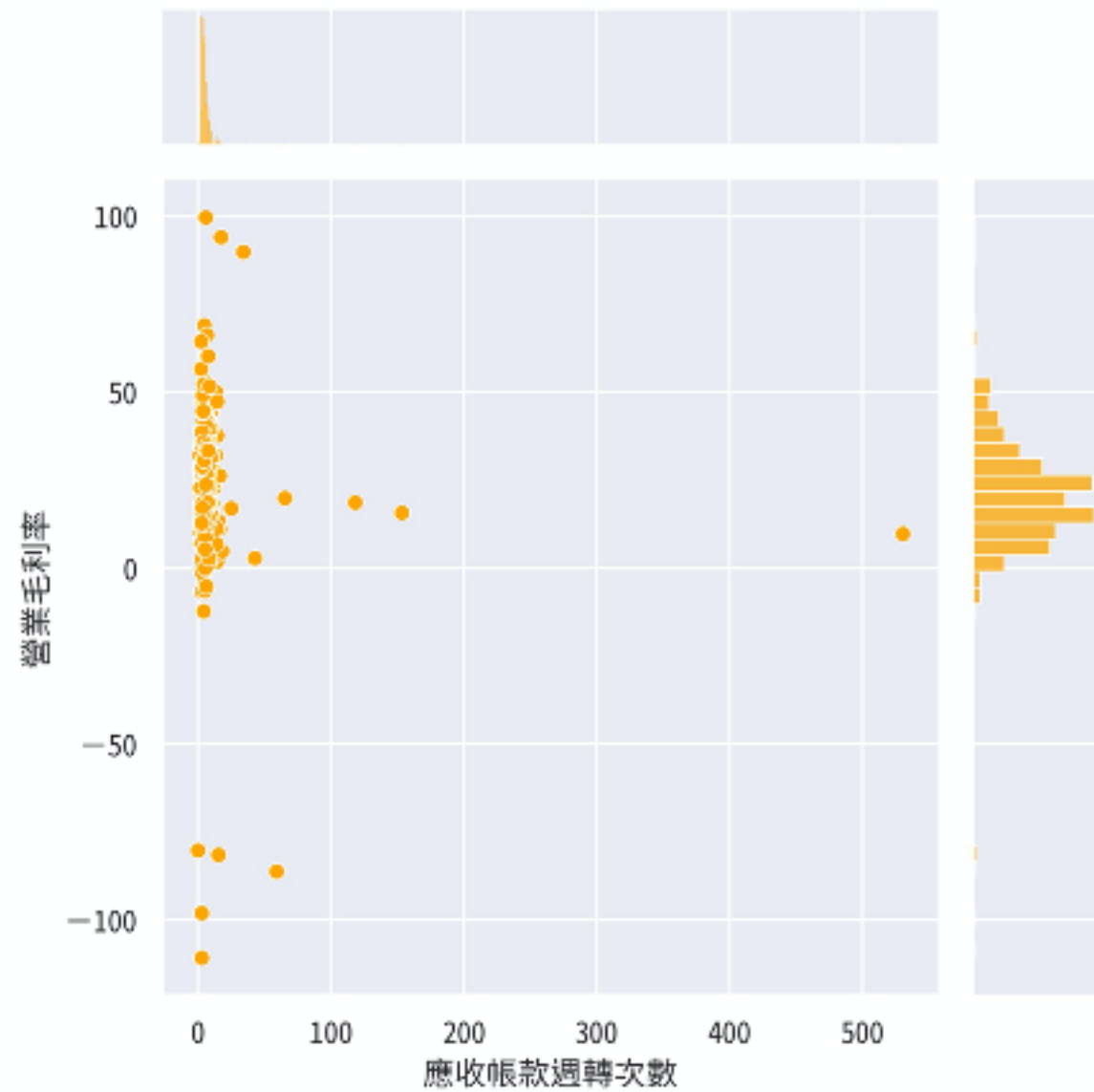
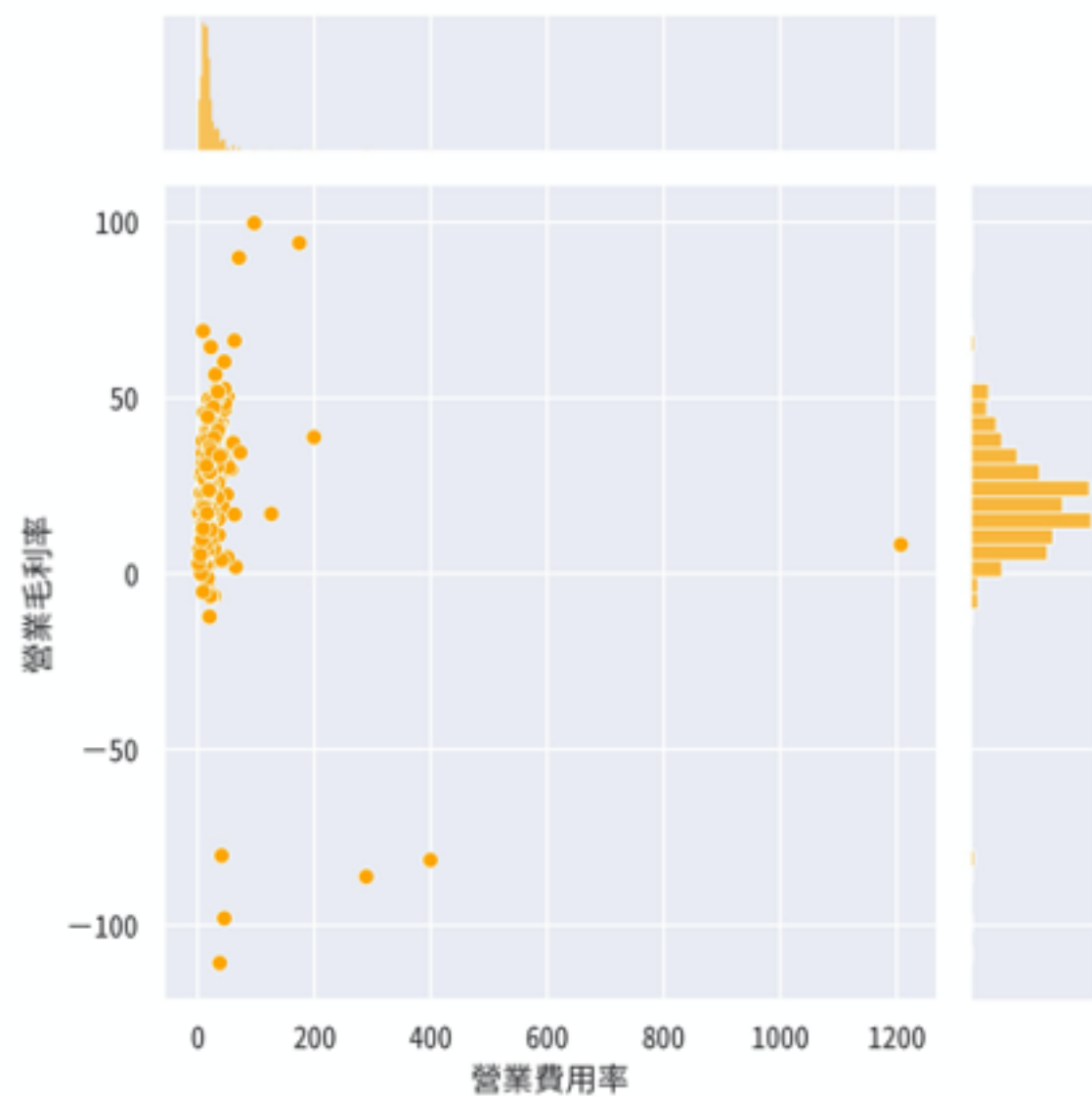
# 營業毛利率和各指標間的關係圖

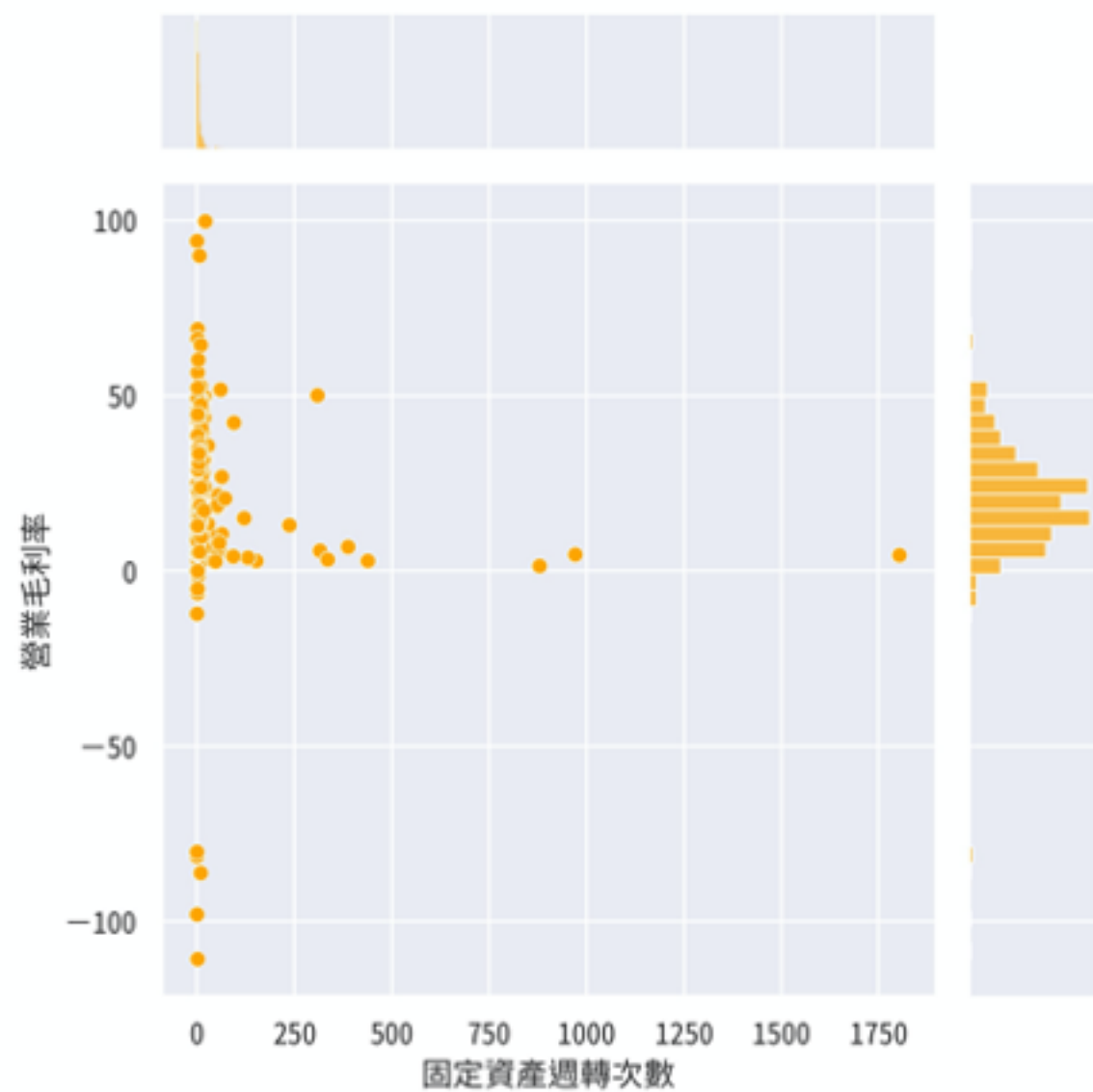
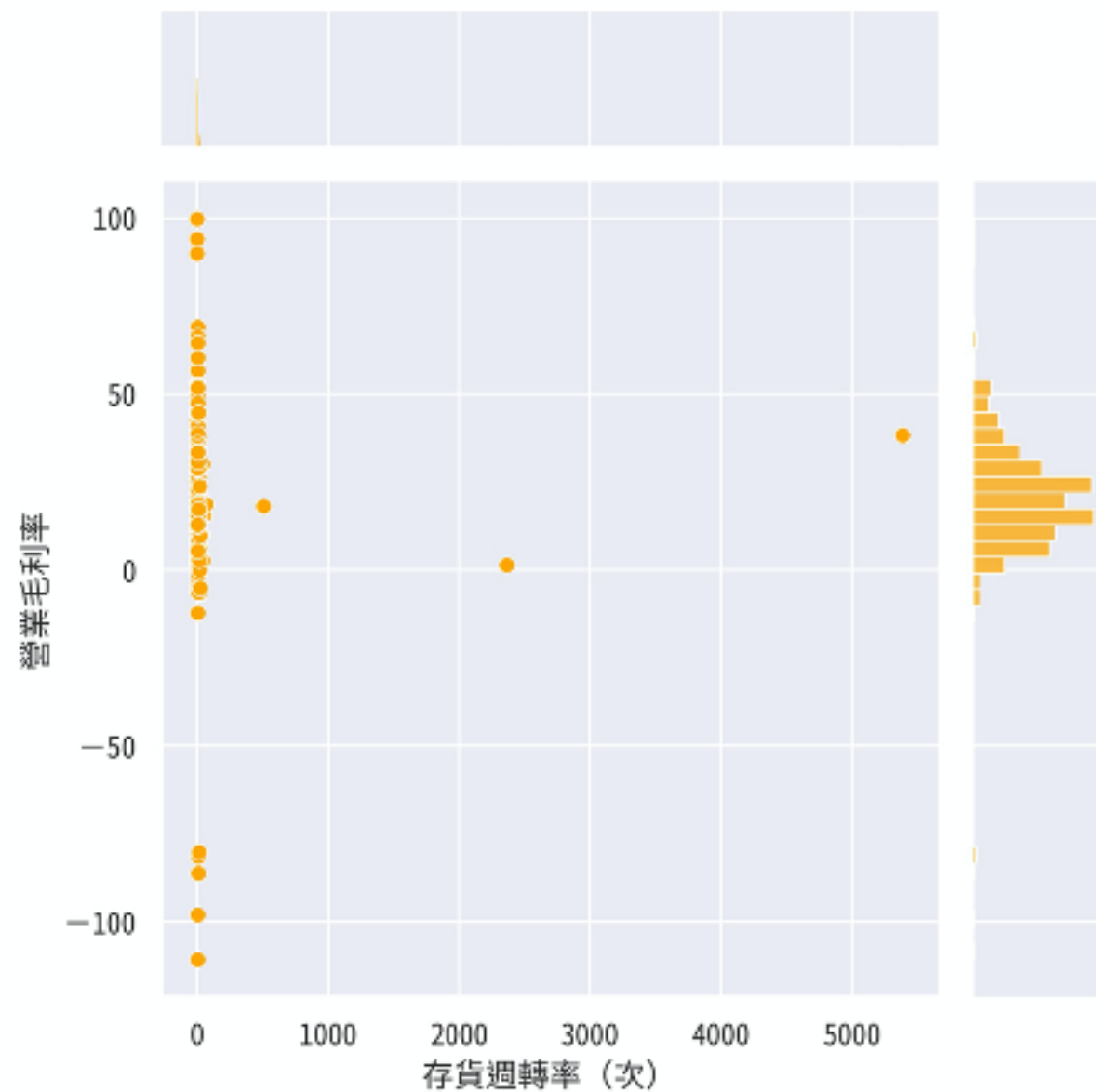


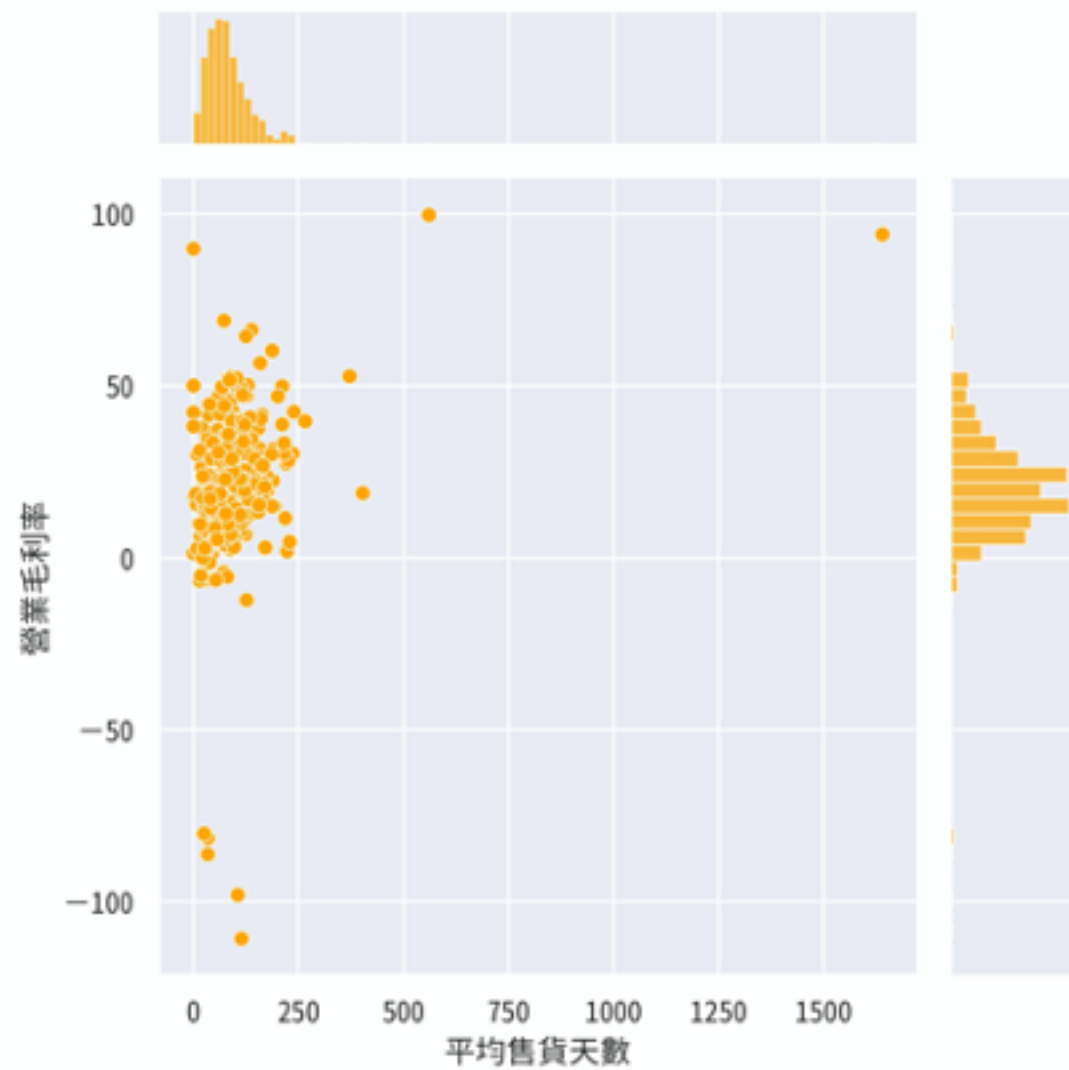
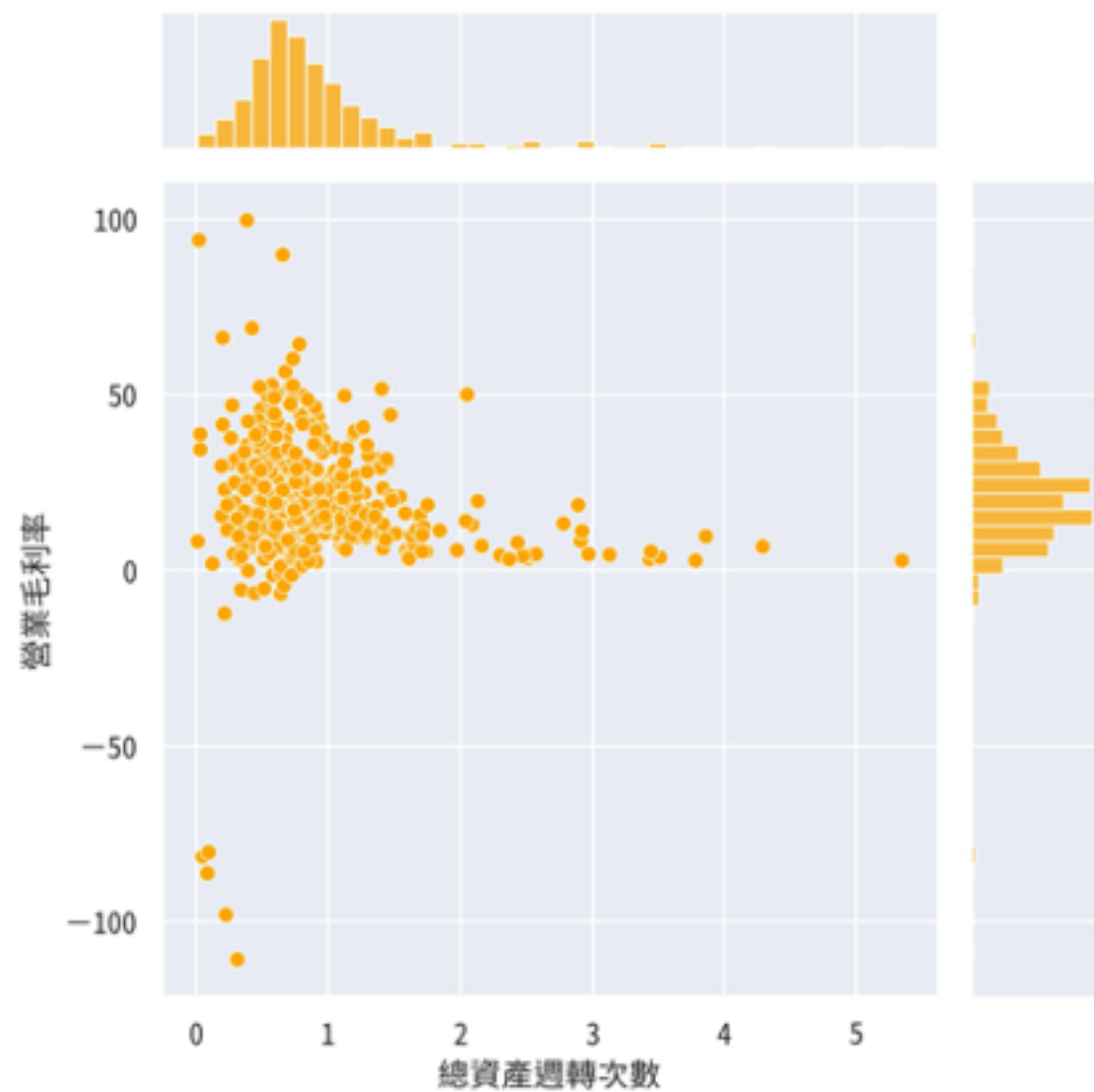




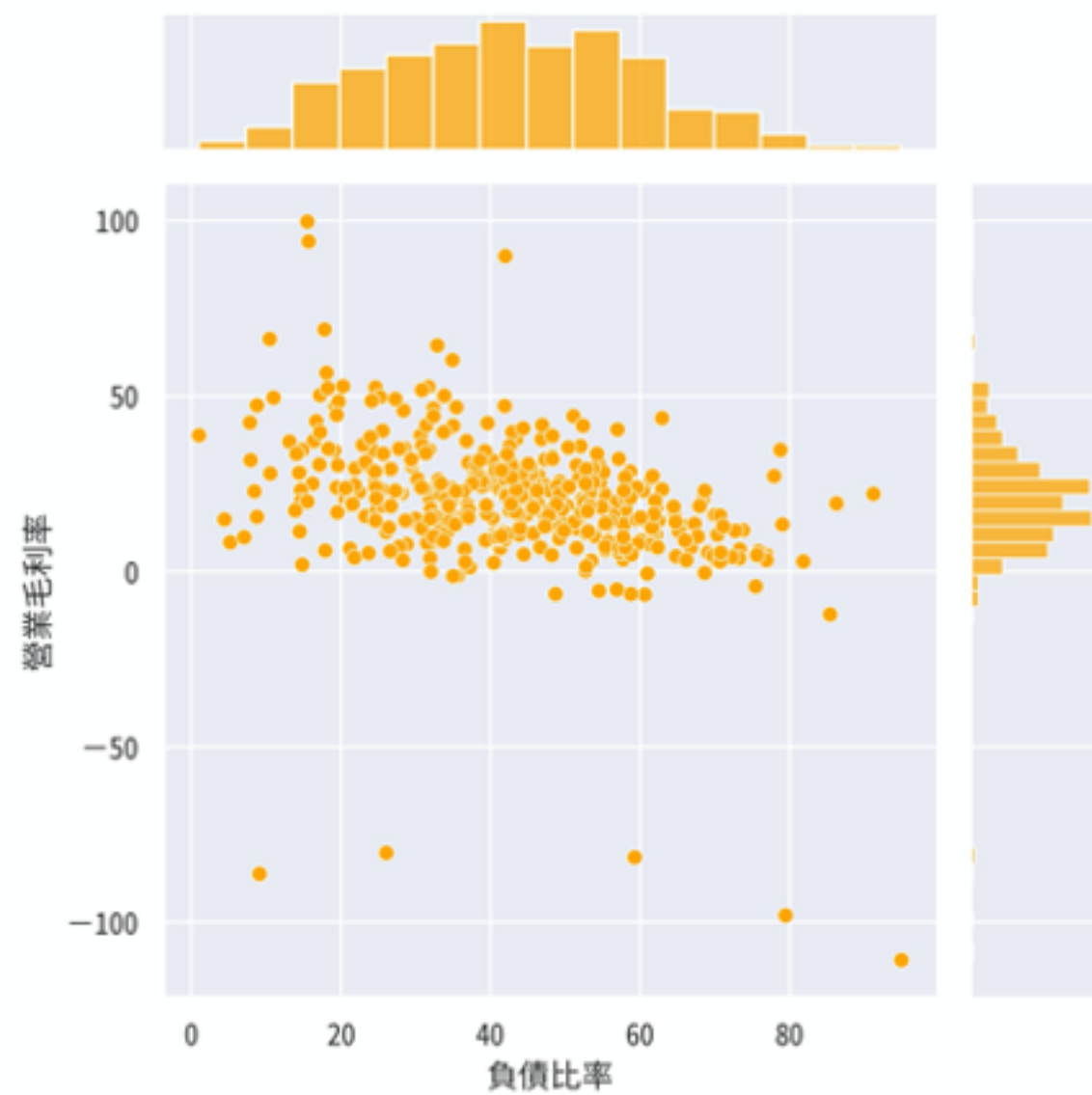
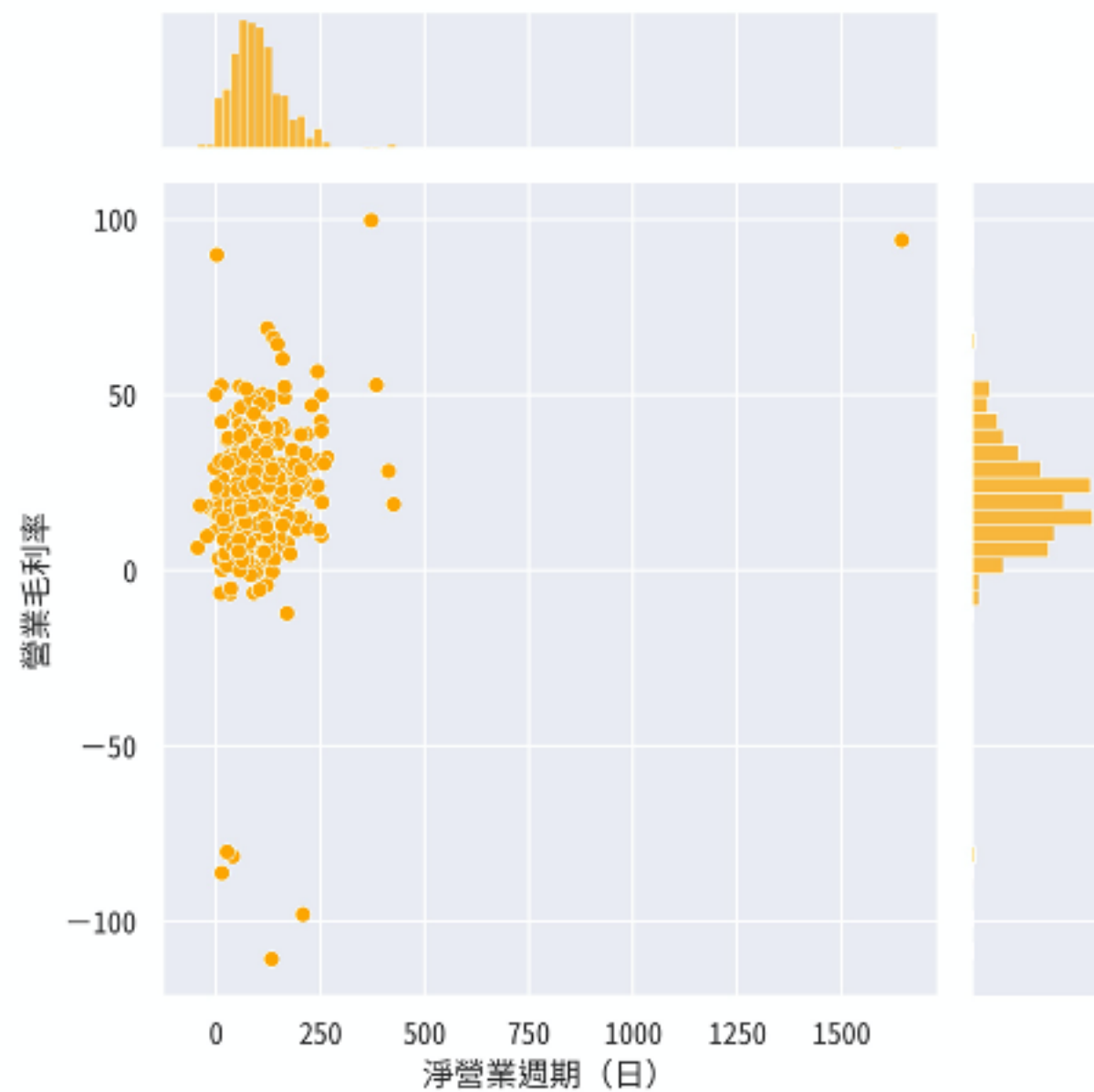


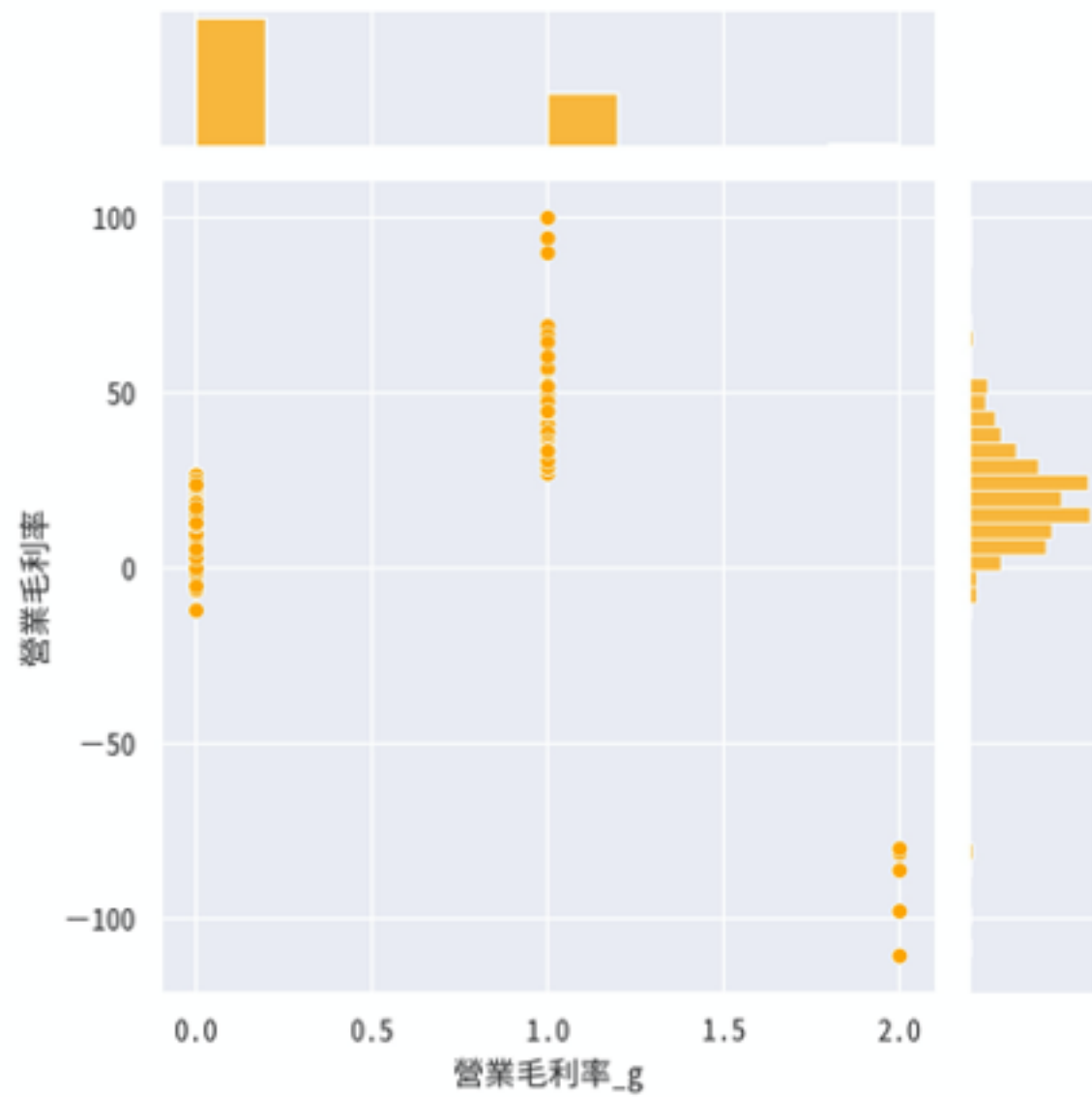
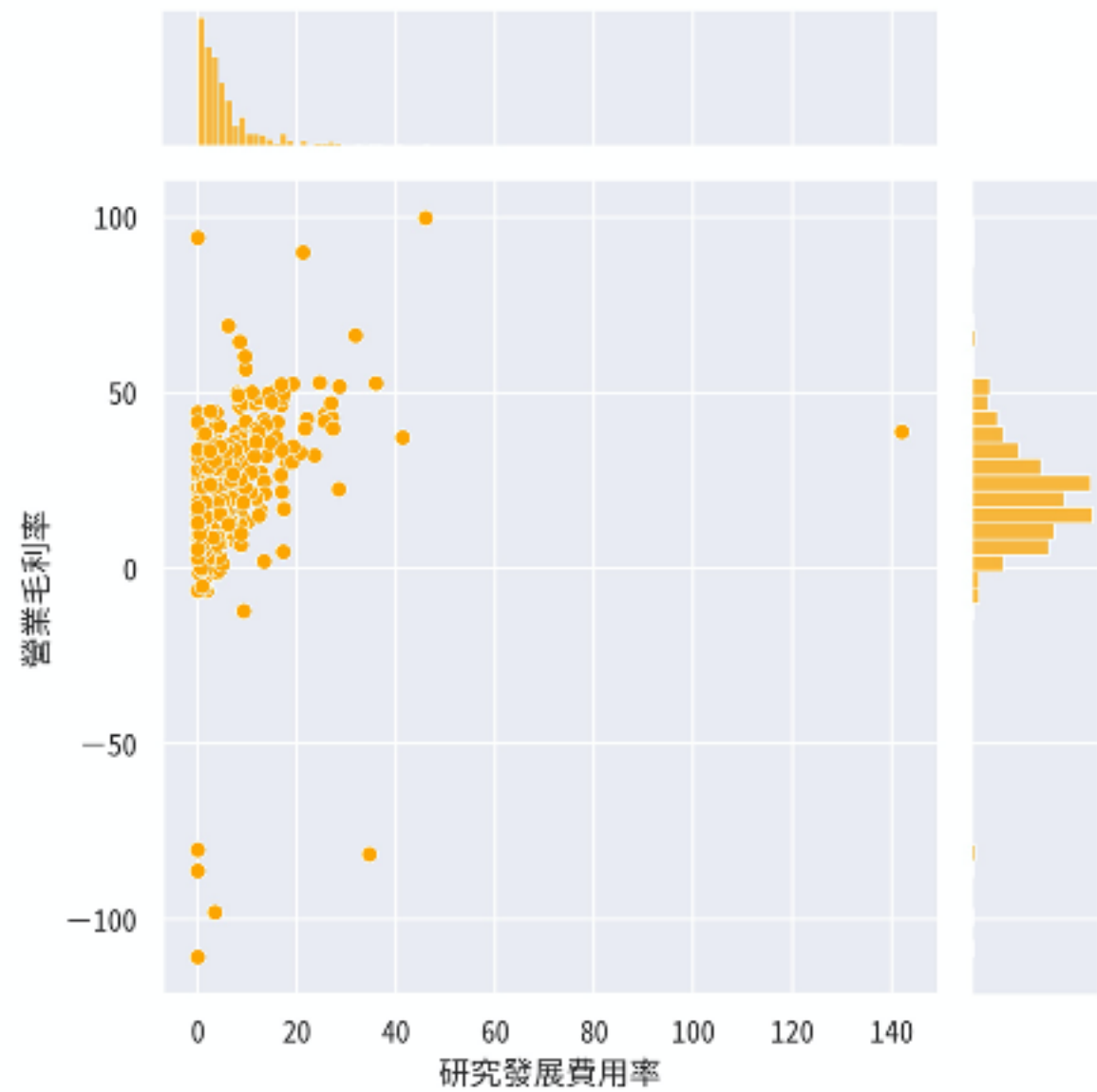








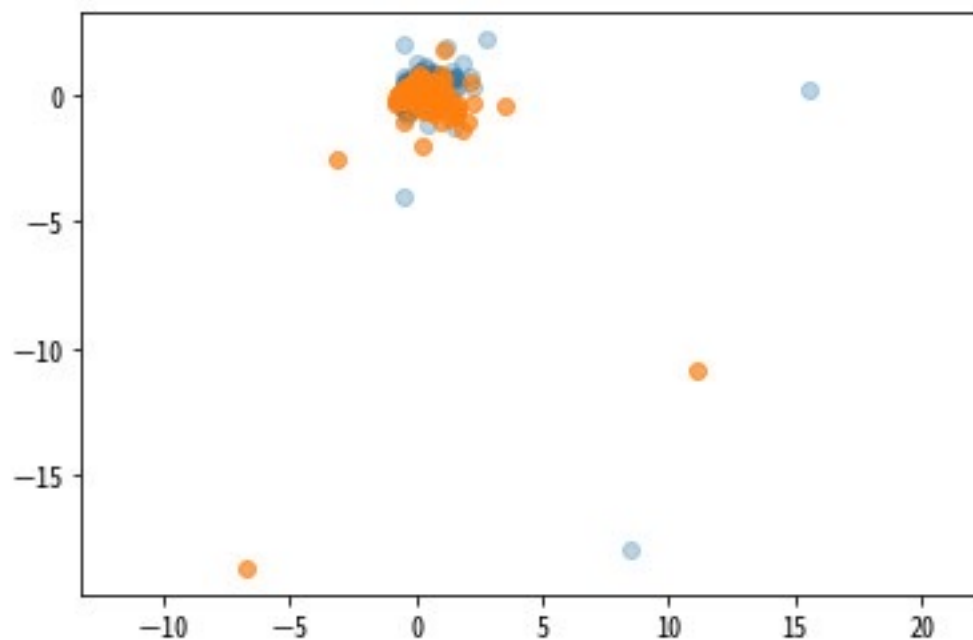






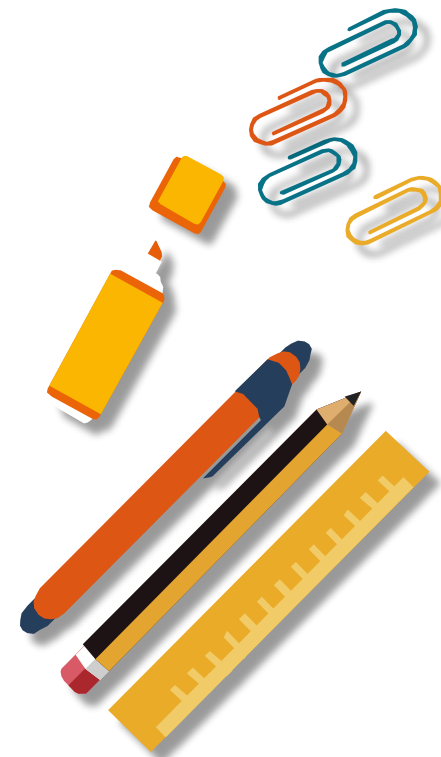
# 各變數分佈小結

Y: 稅後純益成長率



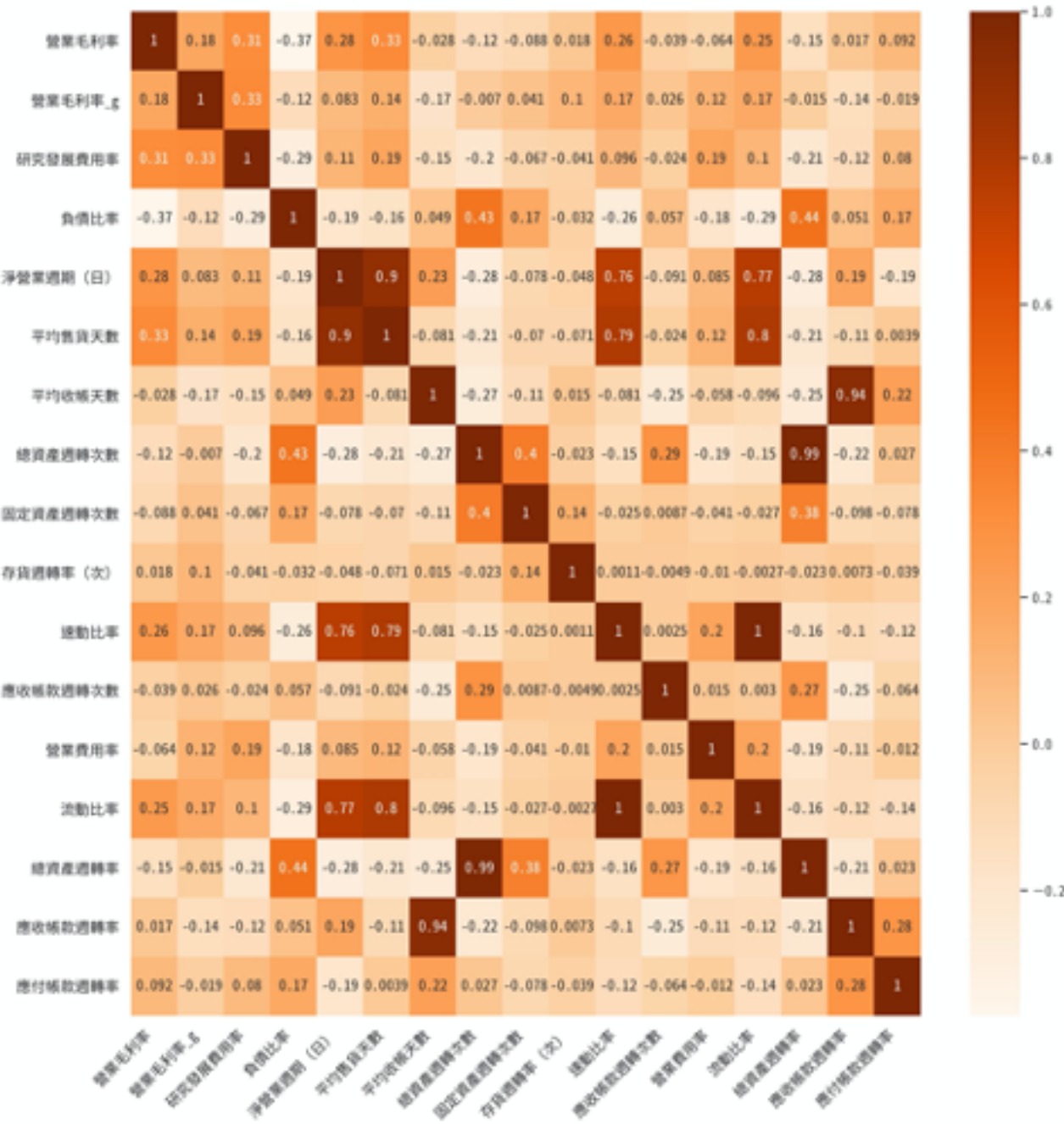
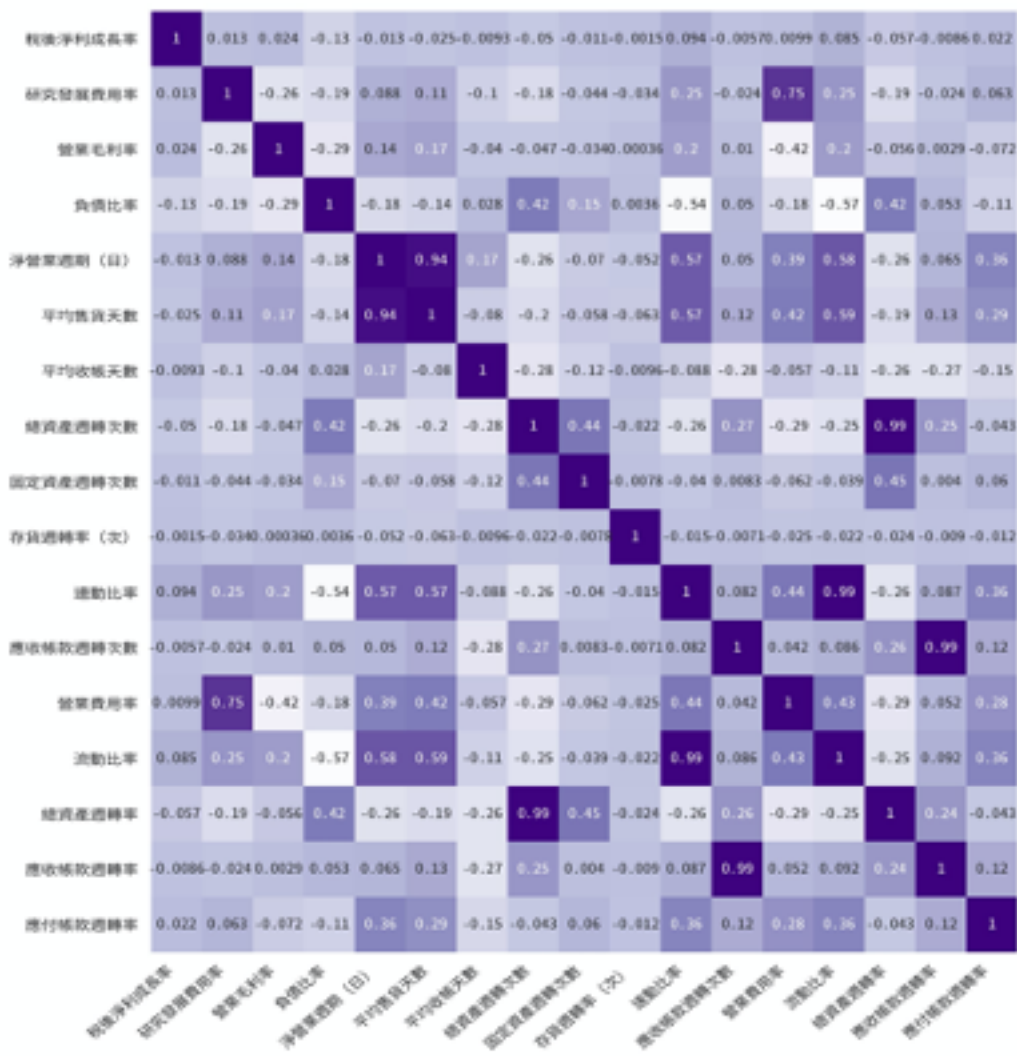
藍色: 原始資料  
黃色: PCA之後

- 每個指標分佈集中  
→ 和挑選過產業有關
- 離散值很少  
→ 特例資料點考慮移除
- 已嘗試PCA  
→ 無改善資料密集狀態, 且可能導致無法解釋





# 各變數間相關係數圖



# 4

## 初步結果說明

- 隨機森林
- 共現性矩陣

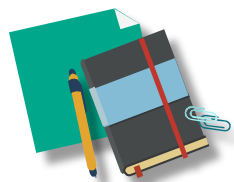


# 隨機森林結果

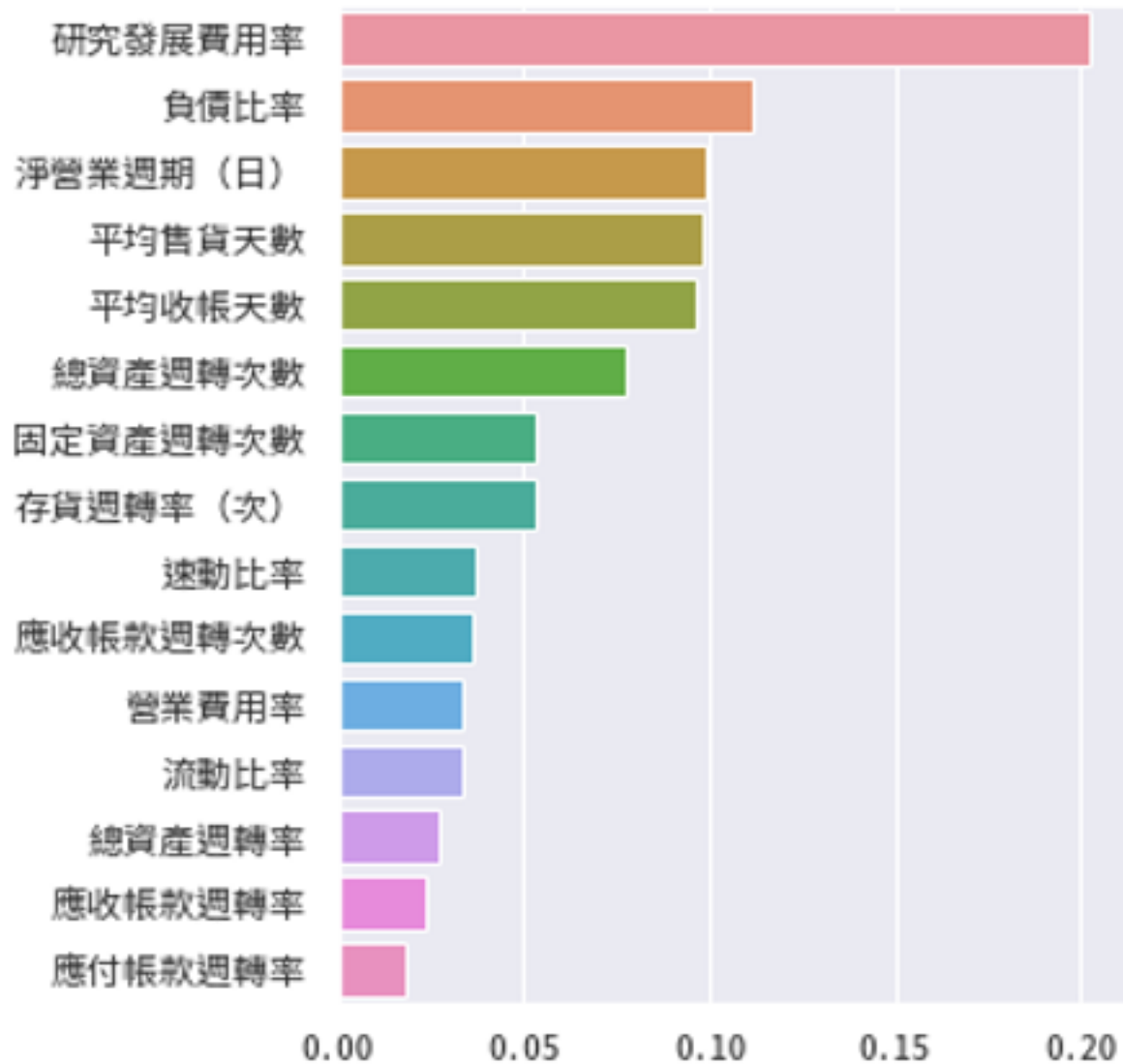


模型準確率：0.8306451612903226

把資料分成10份做交叉驗證模型準確度：0.80 (+/- 0.12)

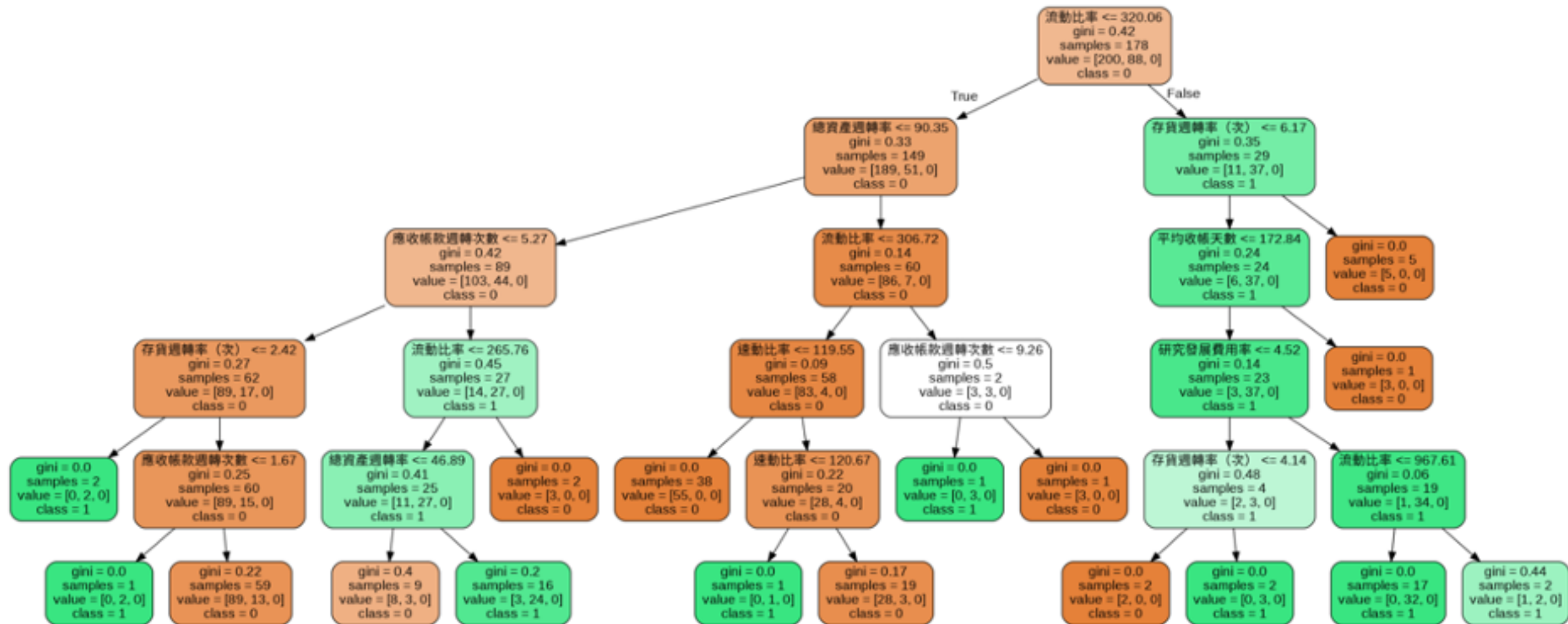


# 指標重要性





# 森林裡其中一棵樹







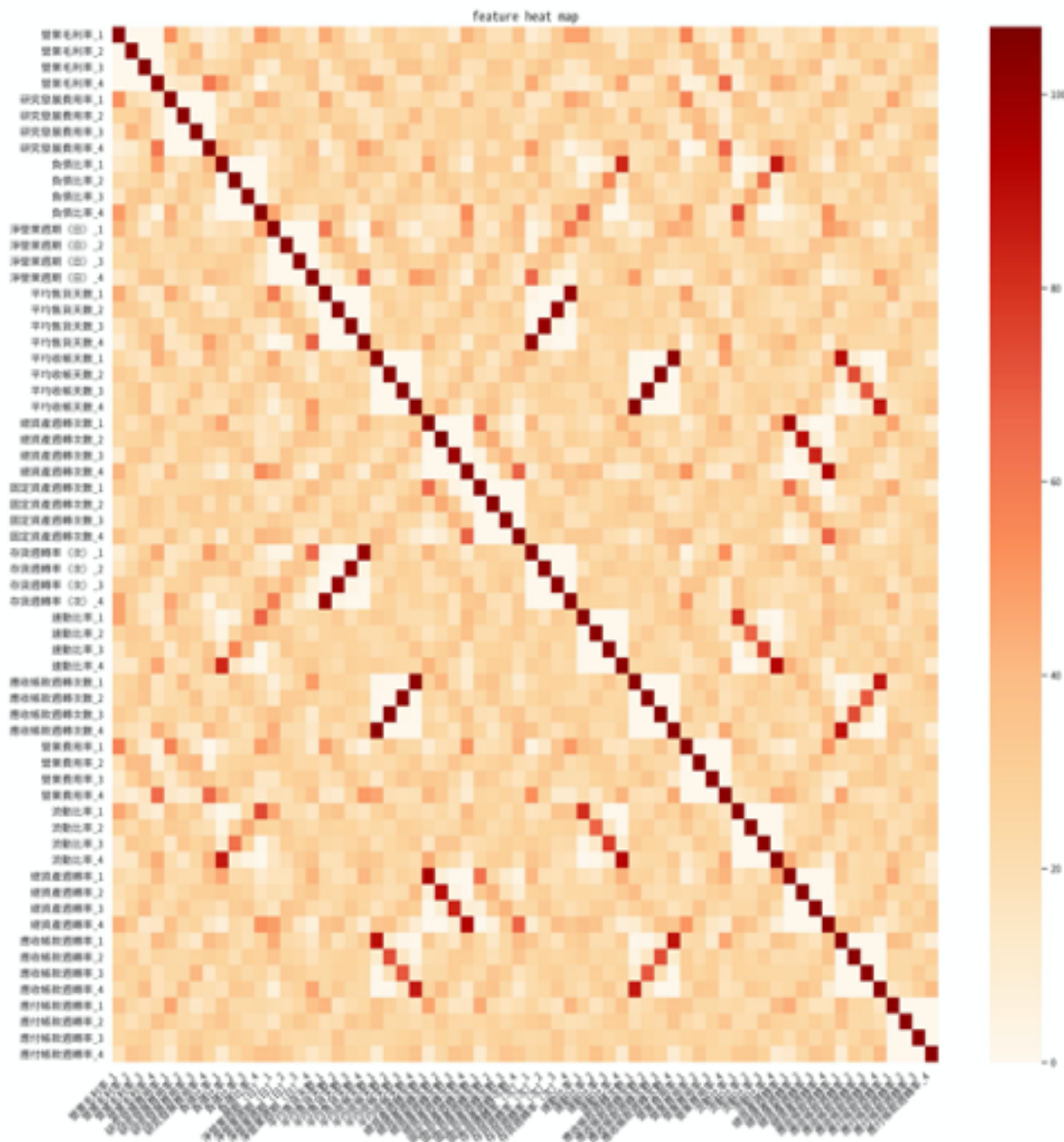
# 共 現 性 矩 陣





# 共線性矩陣

每個特徵以Q1, Q2, Q3  
分成四個區間



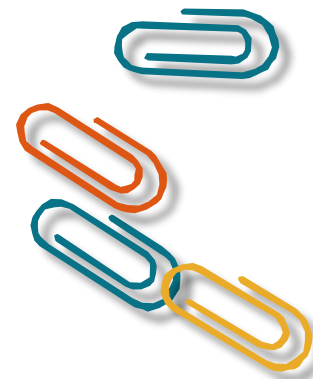
[illegible]

5

## 會議記錄

- 2021/05/19
- 2021/05/26

## 2021/05/19 會議記錄



系所	組員	負責內容
巨資碩一	王怡人	✓ 提出目前問題與主題
巨資碩一	張文綺	✓ 共同討論解決方案
經濟碩一	陳博勁	✓ 指標充新選擇
經濟碩一	張耀仁	✓ 補充參考資料



## 2021/05/26 會議記錄

系所	組員	負責內容
巨資碩一	王怡人	EDM製作
巨資碩一	張文綺	PPT.會議記錄
經濟碩一	陳博勁	TEJ資料整理
經濟碩一	張耀仁	指標構思與解釋



感謝聆聽  
THANKS