

遊戲營收預測

1. [前提概要](#)
2. [模型預測方式 & 評估](#)
3. [模型特徵 & 特徵篩選](#)
4. [模型訓練成效](#)
5. [預測結果](#)
6. [誤差原因調查](#)
7. [實用性評估](#)



營收預測的實際應用

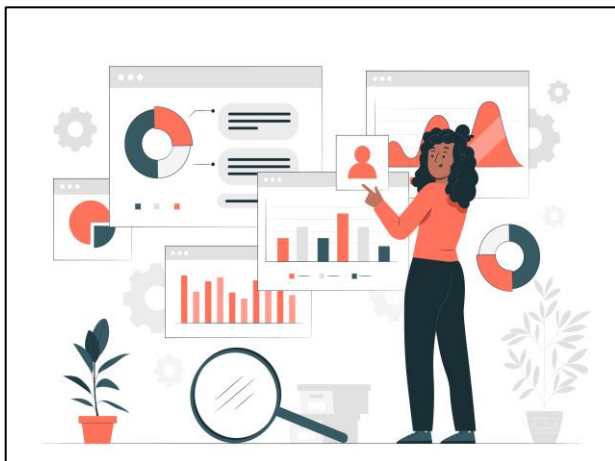


用戶獲取

投放更多資金在高價值渠道上

資源分配

資源資中在高價值用戶體驗上

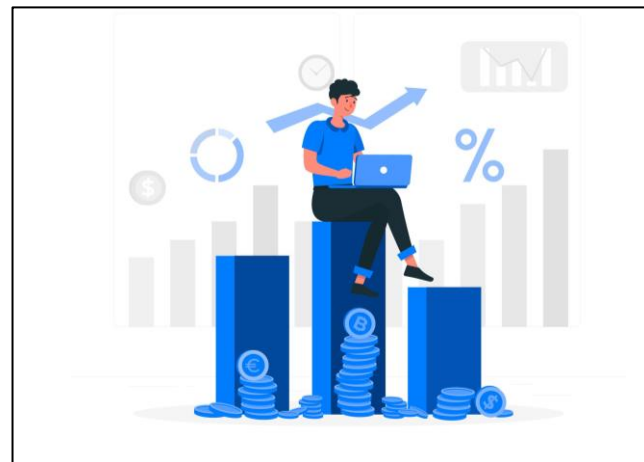


變現&留存

不同玩家量身定製的變現策略

財務預測

準確預測未來收入
評估規劃及預算



前提概要

需求: 每週新用戶180日營收預測

資料需求: 7天

標的: 180日營收預測

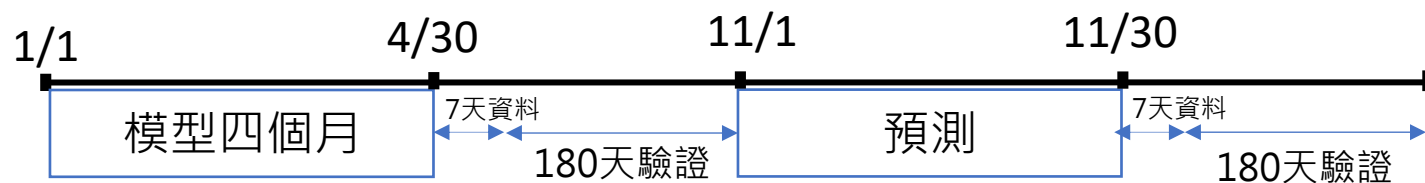
實際應用: 評估當週廣告是否會回本

困難點: 付費率5%，大部分為免費活躍玩家

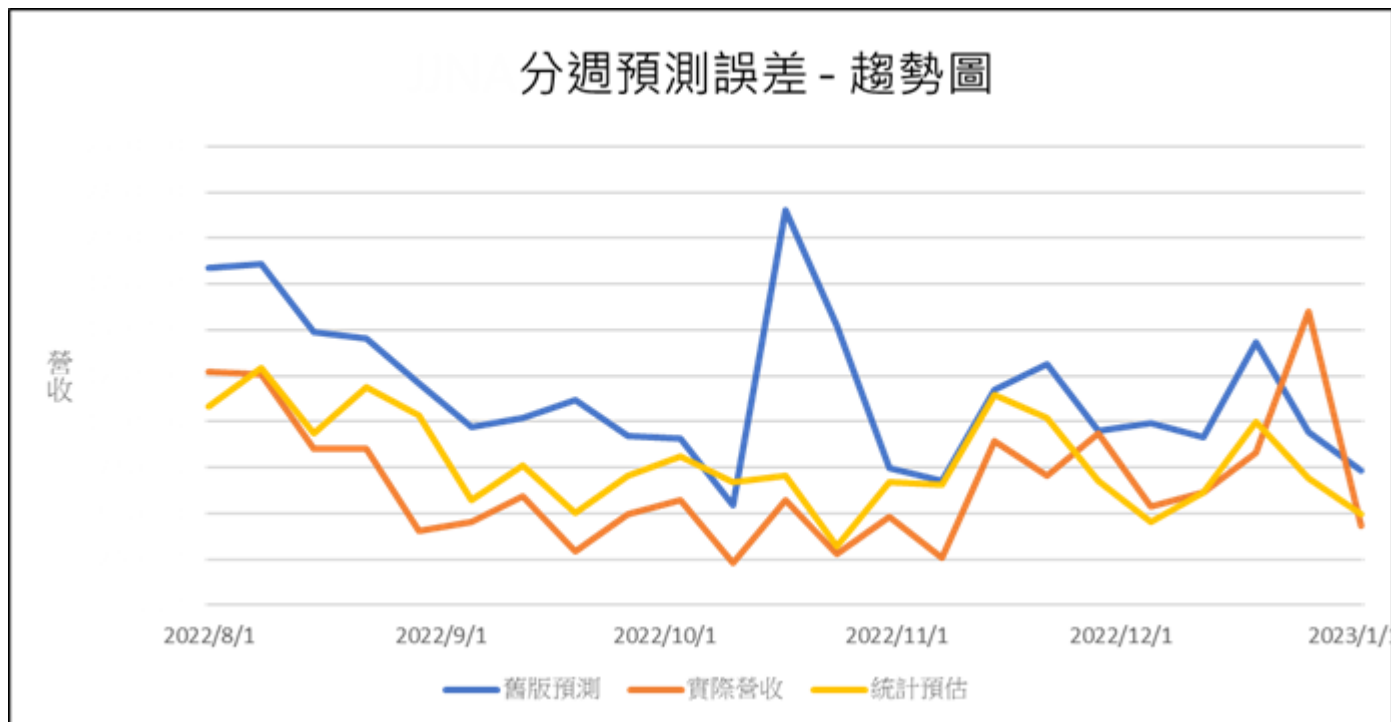
修改:

實際目的是評估回本，如果有需要可以再轉換成 ARPU 或 CPI 的比較維度，此次模型新增多種遊戲
內詳細特徵、優化評估的方式、新增模型

模型預測方式 & 成效評估

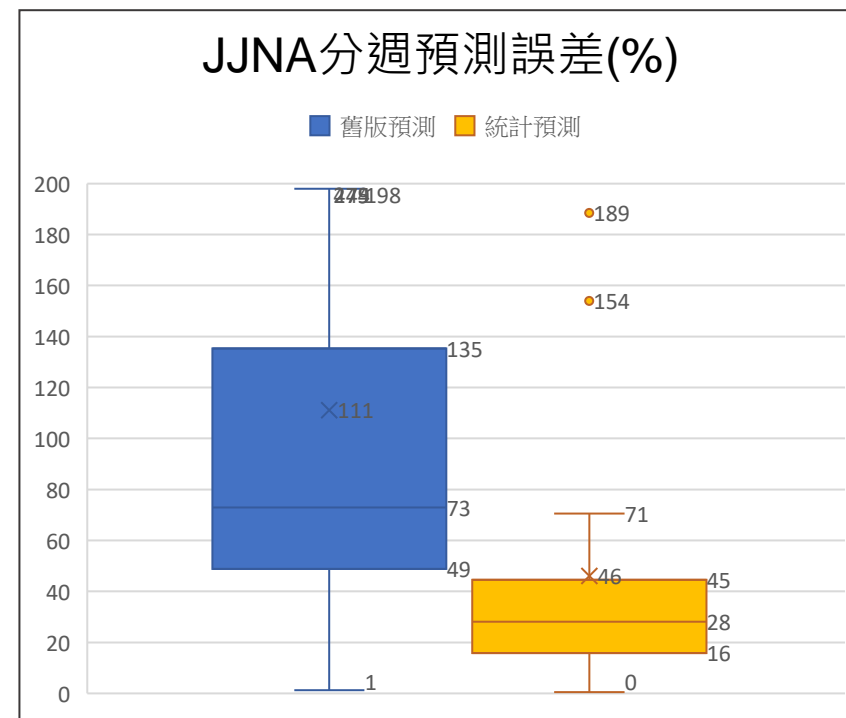


統計預測：預測當週7天儲值 * 訓練期 180天儲值/7天儲值



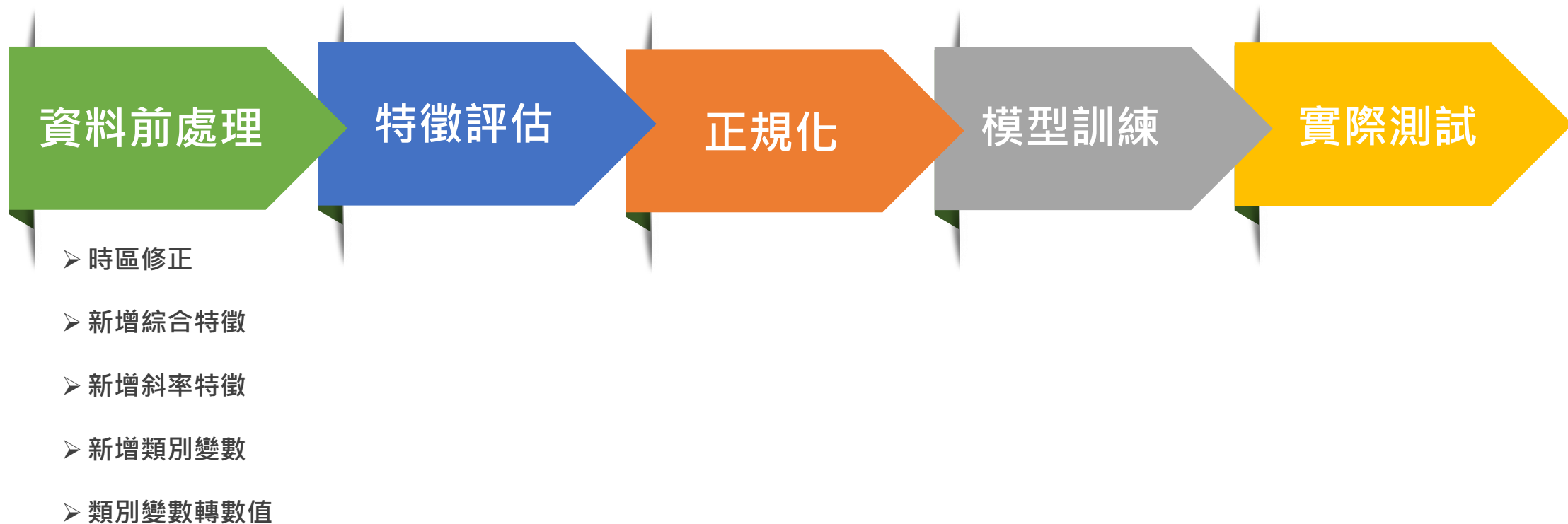
- 舊版預測不準確，統計預測比舊預測準很多

		預測	2022年				
訓練			8月	9月	10月	11月	12月
2021年	10~1月						
	11~2月						
	12~3月						
2022年	1~4月						
	2~5月						
	3~6月						



- 目標是要優於**統計預測**

模型預測步驟



模型特徵

舊版特徵

玩排位賽間隔
玩排位賽次數
玩排位賽勝率
玩一般對戰間隔
玩一般對戰次數
玩一般對戰勝率

上線時間

登入次數

5點後第一次登入時間

最大等級

最大球員等級

是否技能升級

是否有抽任意物品 (球員、卡片、活動)

是否消耗或獲取鑽石

是否完成任務 (日、週、挑戰)

■ 每日數據
■ 整合數據
■ 類別數據



登入、上線類

每日登入
七天登入幾天
最大連續幾天登入
每日上線時間
平均日上線時長
上線時長斜率
上線時長正斜率次數
4點後第一次登入時間斜率
4點後第一次登入時間正斜率次數
第一次登入選擇職業

訓練類

最大訓練階級
順練階級斜率
順練階級斜率正負斜率次數
技能升級總次數
達到技能滿級總數
最大球員評價
最大球員突破

任務類

日任務完成總數
日任務完成斜率
日任務完成正斜率次數

其他

國家
平台

對戰類

對戰次數
對戰平均每天次數
對戰次數斜率
對戰次數斜率次數
對戰總勝率
一般/牌位百分比
一般/牌位類別(70%up)
間隔少於5分鐘場次
間隔少於5分鐘場次 / 間隔場次
平均ping值
是否有真人對戰
真人對戰場次
真人對戰場次/總場次
真人對戰勝率
真人對戰評價差異
對戰評價差異

儲值、廣告類

首儲日
首儲金額
7日總付費金額
7日總儲值次數
平均每日儲值
儲值金額斜率
儲值金額正負斜率加總
首點廣告日
7日總廣告金額
平均每日廣告點擊次數
廣告點擊斜率
廣告點擊正斜率次數

一次性鑽石消費

技能槽 (1個、2個)
月卡 (7天、30天)
戰令通行證(普通、進階)
日常戰令

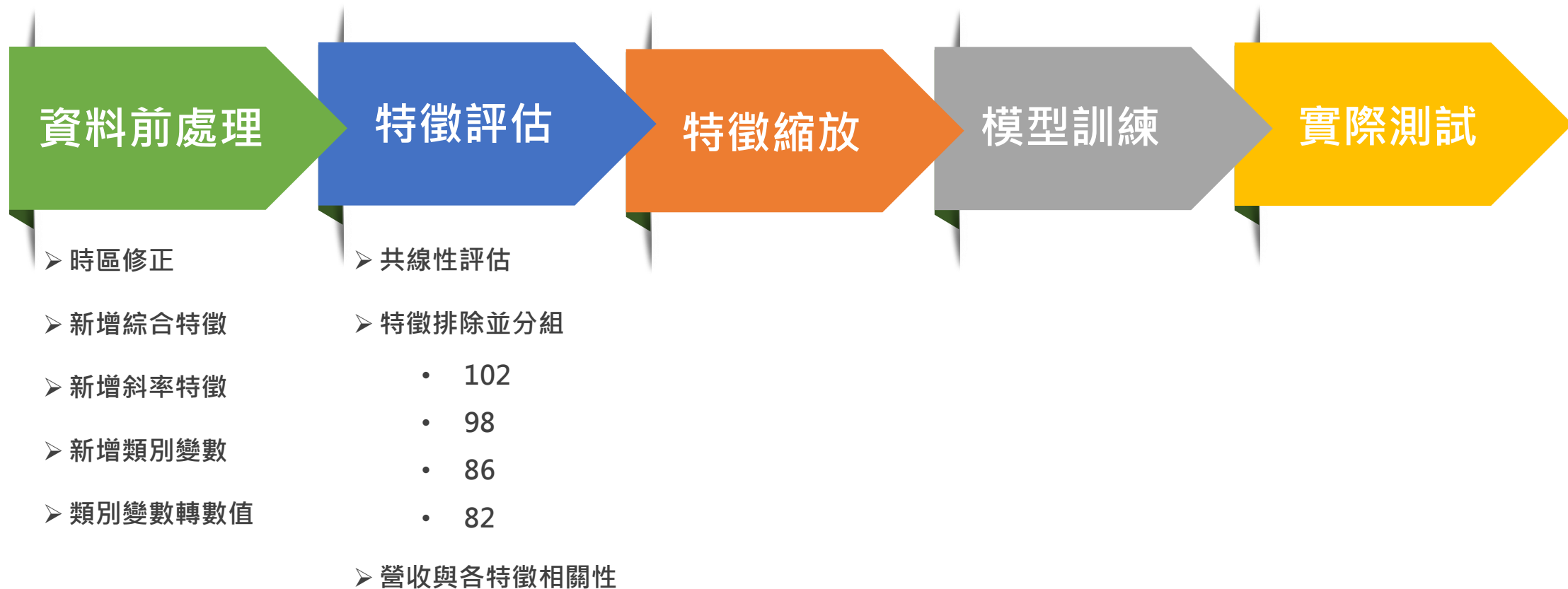
鑽石總消費

球員金幣、銀幣
金星、銀星
鑽石轉盤
購買球員
永久商店限購購買
球員簽約商店購買
購買服裝
戰令商店經驗購買
球員突破商店購買
花式動作商店購買
菁英卡牌商店
時裝選擇屬性
兌換金幣

綜合性鑽石消費

鑽石消費
總消耗鑽石
平均每日消耗鑽石
鑽石消耗斜率
鑽石消耗正負斜率次數
總消耗綁鑽
平均每日消耗綁鑽
綁鑽消耗斜率
綁鑽消耗正負斜率次數

模型預測步驟



什麼是共線性？

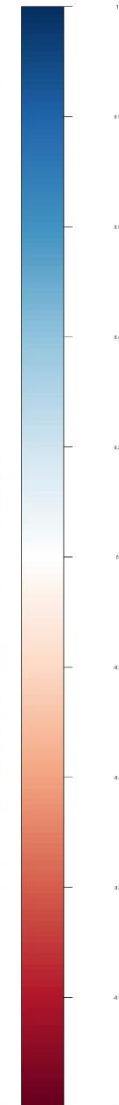


我正在研發我的饅頭
想找出哪種饅頭最好吃

1. 麵粉重量
2. 鹽巴
3. 糖
4. 大小
5. 顏色
6. 香味
7. 甜味

使用的糖跟甜味相關性高，兩個特徵
同時放入模型會有共線性的情形

各特徵相關性矩正

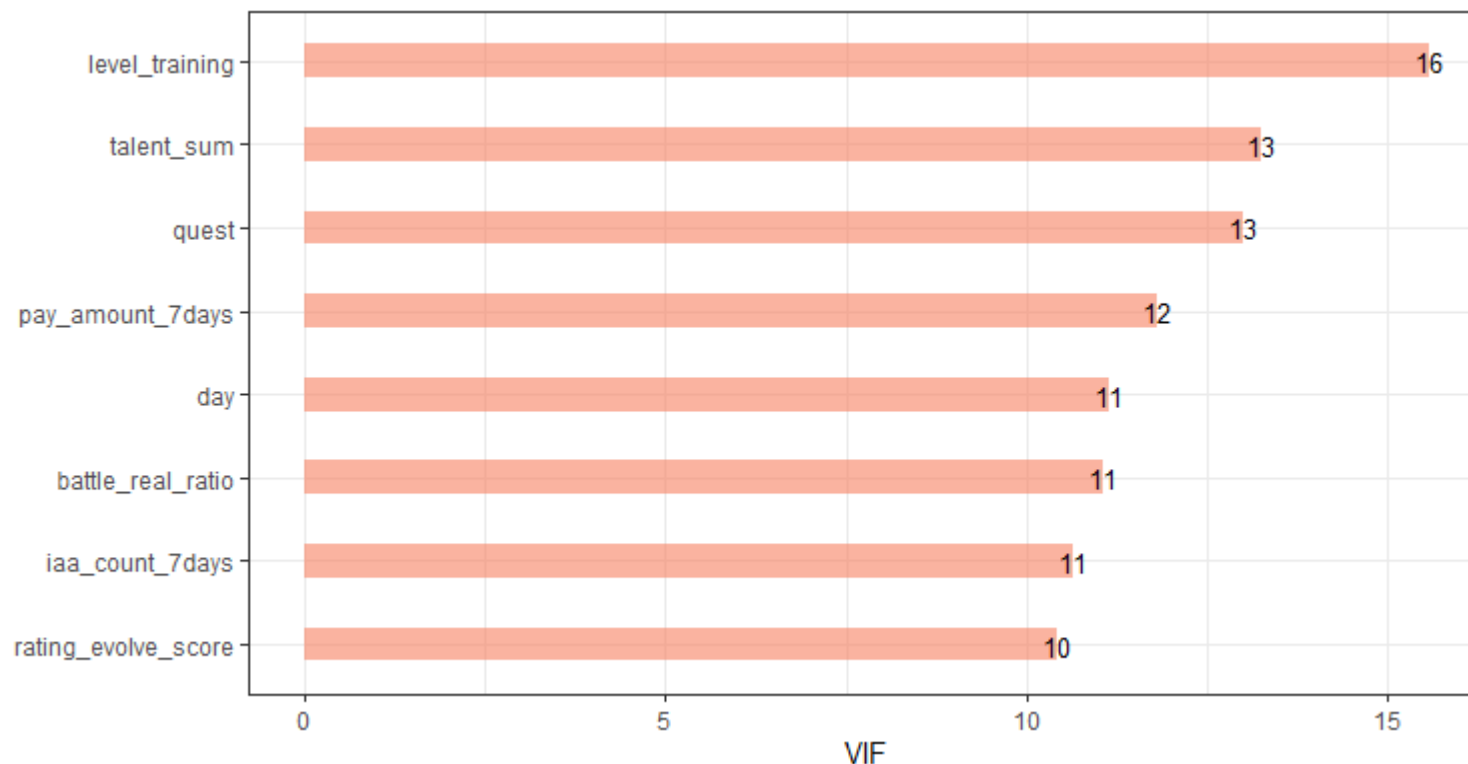


排除後VIF共線性(3/3)

共線性: $VIF > 10$

Condition index: 38.48
(超過30代表在線性模型下有共線性)

VIF Factor



- 排除21個特徵後剩餘86個特徵，部分特徵還是有些微的共線性
- 之後使用的模型都可以一定程度消除共線性，所以不一定要把全部排除
- 共線性高到低，主觀選擇4種特徵組合: 102個、98個、86個、82個

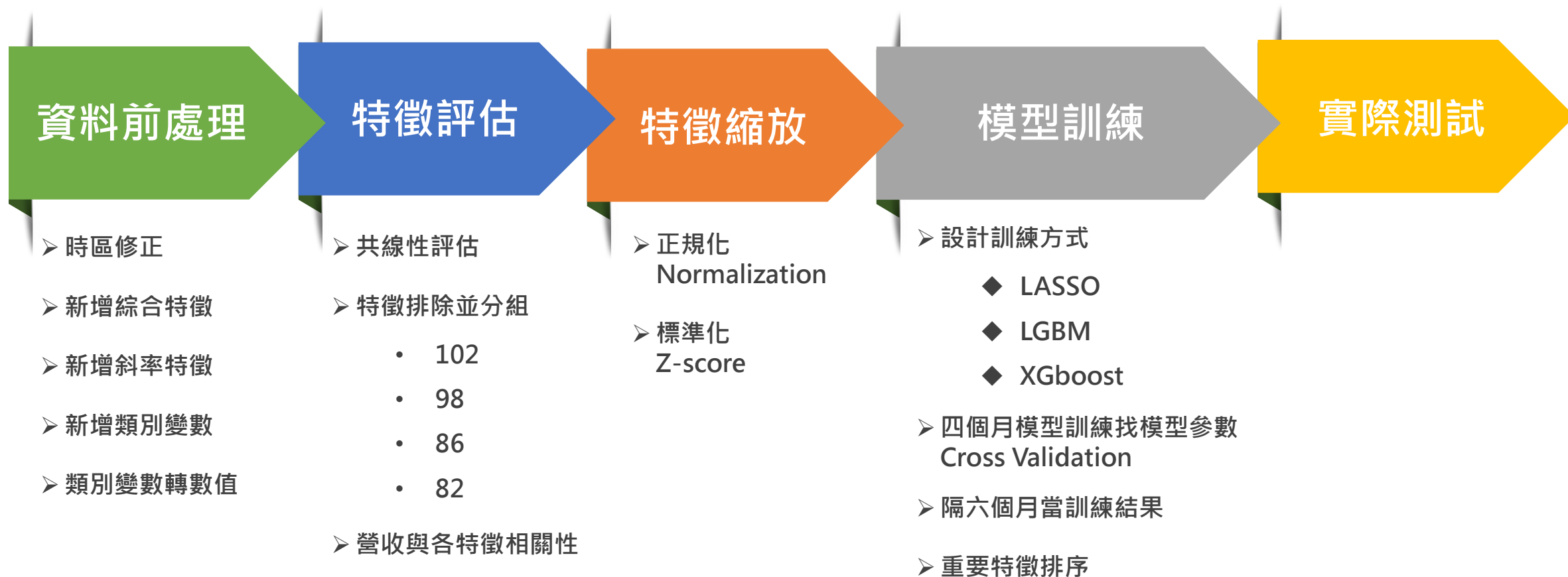
累積180日營收與特徵相關性

7日總付費
7日總付費次數
每日平均付費
簽約球員商店
第一次付費金額
限購商店
戰令
球員金幣
技能欄
月卡

Index	revenue
revenue	1
pay_amount_7days	0.453983
pay_count_7days	0.366973
pay_amount_avg	0.339509
diamond_contract_player_shop	0.308578
first_pay_amount	0.308187
diamond_limit_shop	0.304827
diamond_battle_pass	0.294055
diamond_player_coin	0.279787
diamond_skill_slot	0.267284
diamond_monthly_card	0.255226

sapphire_consume	0.254437
diamond_daily_pass	0.234505
diamond_buy_player	0.229315
diamond_clothes_status	0.223689
diamond_evolve_shop	0.222502
sapphire_consume_avg	0.21146
diamond_player_coin_avg	0.201901
diamond_contract_player_shop_avg	0.198553
diamond_player_star	0.198267
battle_real_count	0.189258
diamond_limit_shop_avg	0.188748
diamond_player_star_avg	0.179903
onlnte_time_day6	0.175381
first_pay_day	0.17353
diamond_clothes_status_avg	0.171411
onlnte_time_day7	0.171178
iaa_count_7days	0.170866

模型預測步驟



各模型訓練及各特徵數下誤差

Train: 2022/1~4月
Test: 2022/11月

2022年11月	180日累積營收	預測誤差
實際	250	-
統計預測	350	34.6%
舊預測	400	41.1%

Test result:

跑50次10-fold Cross Validation 找最佳參數

MM = Min-Max normalization

	103	103MM	98	98MM	86	86MM	82	82MM
LASSO_CV	23.26	28.71	22.49	28.71	23.99	29.13	19.3	26.75
LASSO_CV_50	23.23	32.91	23.24	32.05	23.1	31.28	19.51	31.29
lgbm_short	26.14	30.41	22.44	25.02	27.48	25.06	28.63	31.11
lgbm_long	27.64	29.88	28.63	23.55	33.04	27.38	20.72	25.88
xgboost_short	21.27	28.86	29.4	30.03	33.2	22.12	26.34	20.81
xgboost_long	19.19	27.61	28.27	31.09	39.85	24.01	21.34	29.91

重要特徵排序

LASSO

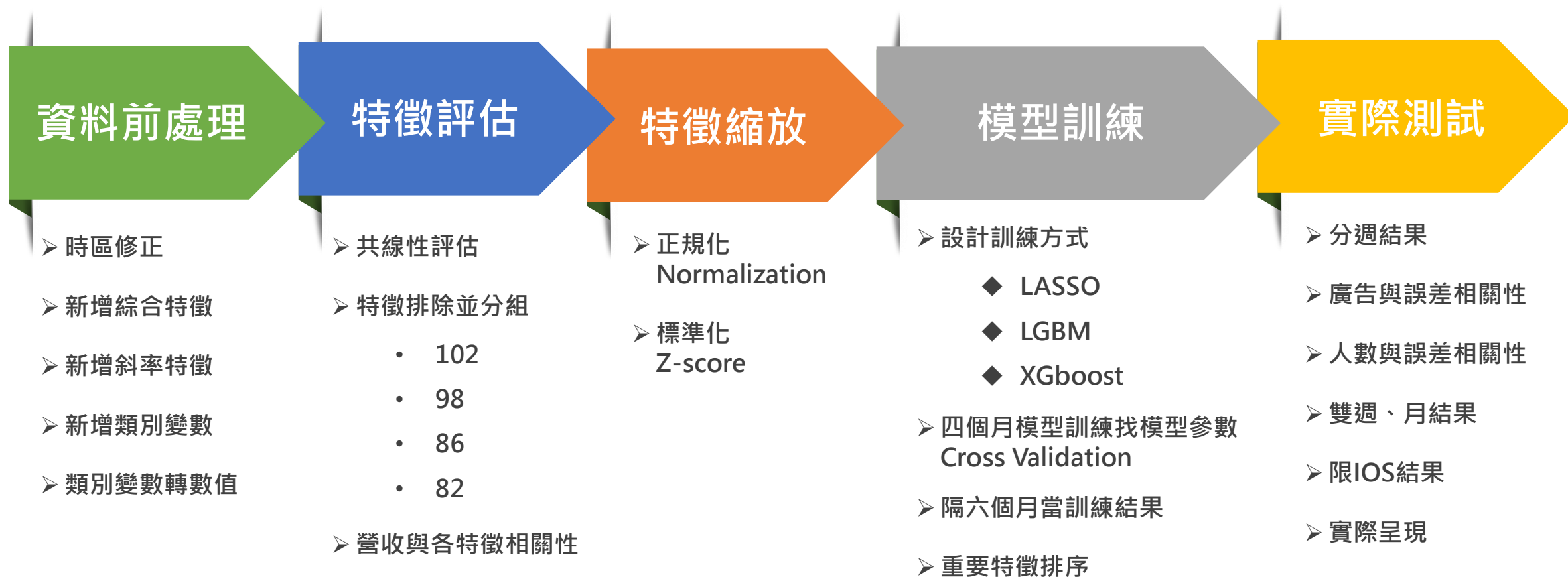
Features	特徵	coefficients
pay_amount_7days	7日總儲值	0.6876
diamond_limit_shop	限購商店	0.1402
diamond_clothes_status	時裝屬性	0.1042
diamond_battle_pass	戰令	0.0801
diamond_contract_player_shop	簽約球員商店	0.0769
diamond_monthly_card	月卡	0.0644
diamond_buy_player	球員購買	0.0590
diamond_player_star	金星、銀星	0.0528
diamond_skill_slot	技能槽	0.0512
diamond_player_coin	球員金幣	0.0369
diamond_evolve_shop		0.0344
sapphire_consume_avg		0.0105
diamond_consume_slope		0.0101
sapphire_consume_slope		0.0024
diamond_clothes		0.0010
onlinte_time_day7		0.0010
onlinte_time_day6		0.0009
onlinte_time_day5		0.0004
onlinte_time_day3		0.0002
onlinte_time_day2		0.0001

LGBM

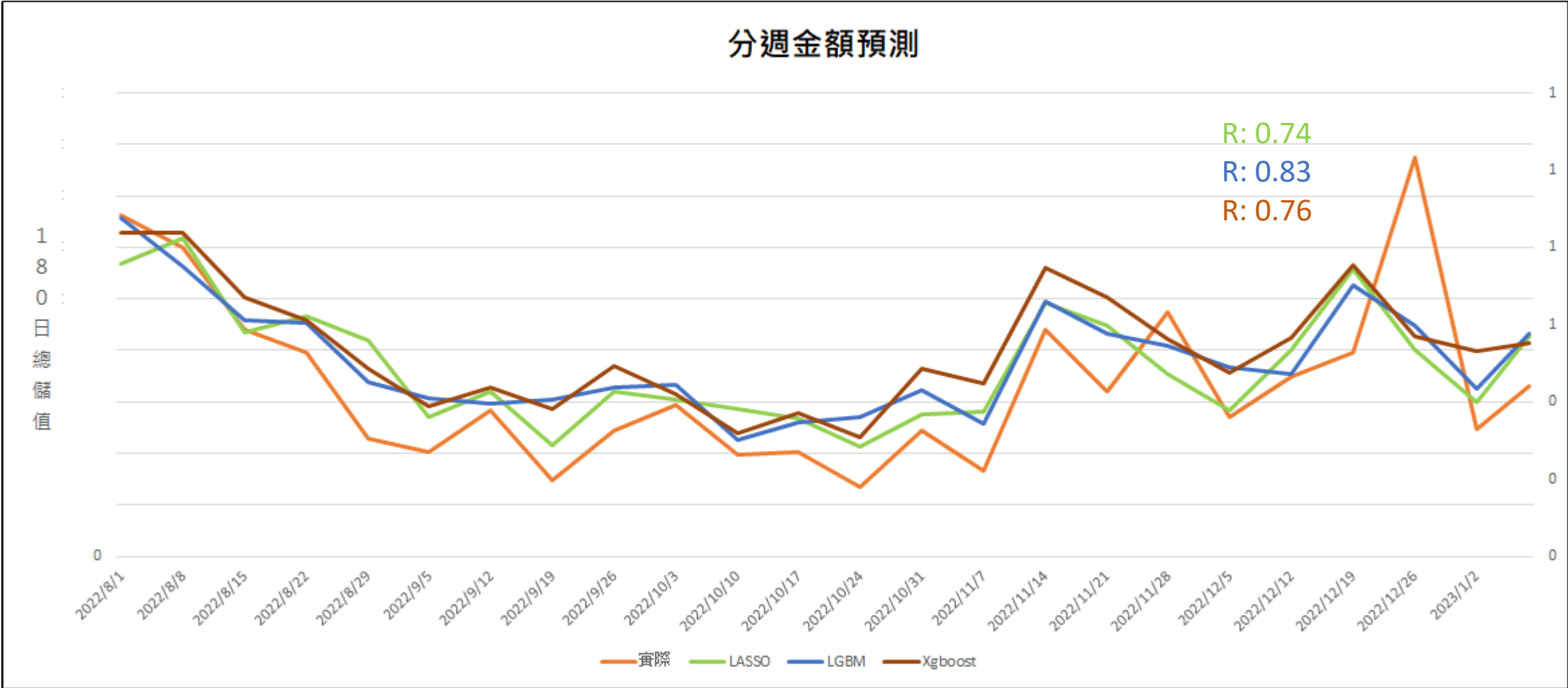
Feature	特徵	Importance
pay_amount_avg	平均每日儲值	494
pay_amount_7days	7日總儲值	395
iaa_income_7days	7日總IAA	208
sapphire_consume_avg	平均每日綁鑽	195
quest	每日任務總數	168
iaa_count_7days	7日總IAA點擊	137
onlinte_time_day3	第三天上線時長	135
pay_count_7days	7日總付費次數	130
onlinte_time_day7	第七天上限時長	124
level_training	最大訓練等級	112
talent_sum		112
first_pay_amount		108
first_pay_day		98
diamond_battle_pass		95
onlinte_time_day6		82
max_rating		79
diamond_buy_player		73
battle_real_count		73
diamond_limit_shop		71
diamond_consume_slope		67

(選取28個特徵)

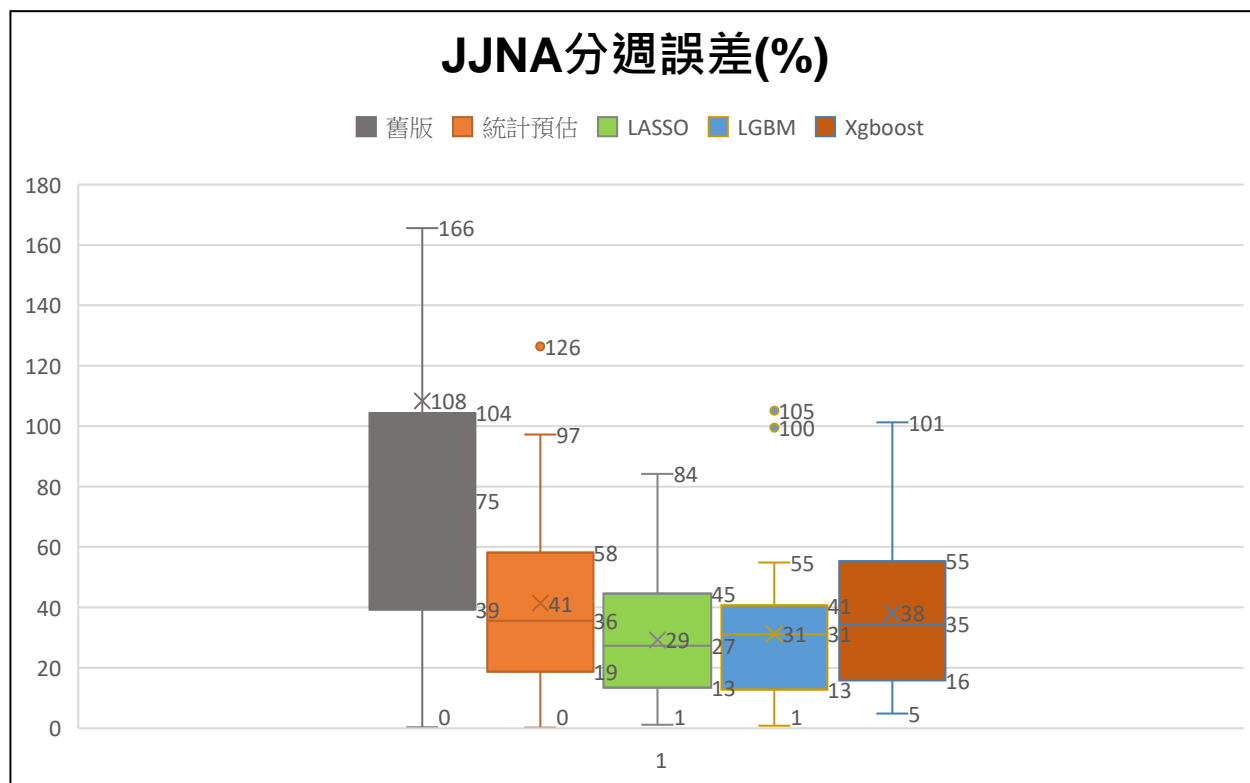
模型預測步驟



分週180日金額預測 – 趨勢圖



分週誤差－盒鬚圖



- 訓練的模型都比統計的結果好不少
- LASSO的平均跟中位數都比LGBM低，且極端誤差較少
- LASSO的結果
 - 相關性R值0.74
 - 25%資料誤差在13%
 - 50%資料誤差在27%
 - 75%資料誤差在45%
- LGBM的結果
 - 相關性R值0.83
 - 25%資料誤差在13%
 - 50%資料誤差在31%
 - 75%資料誤差在41%
- Xgboost結果差

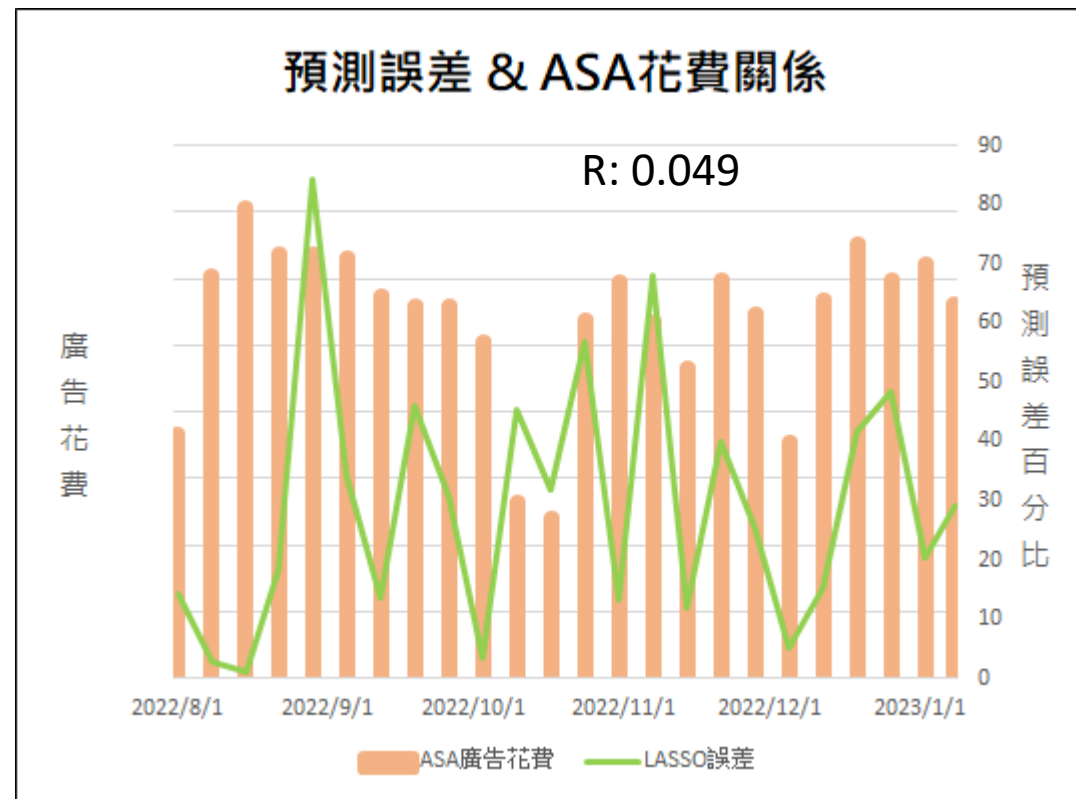
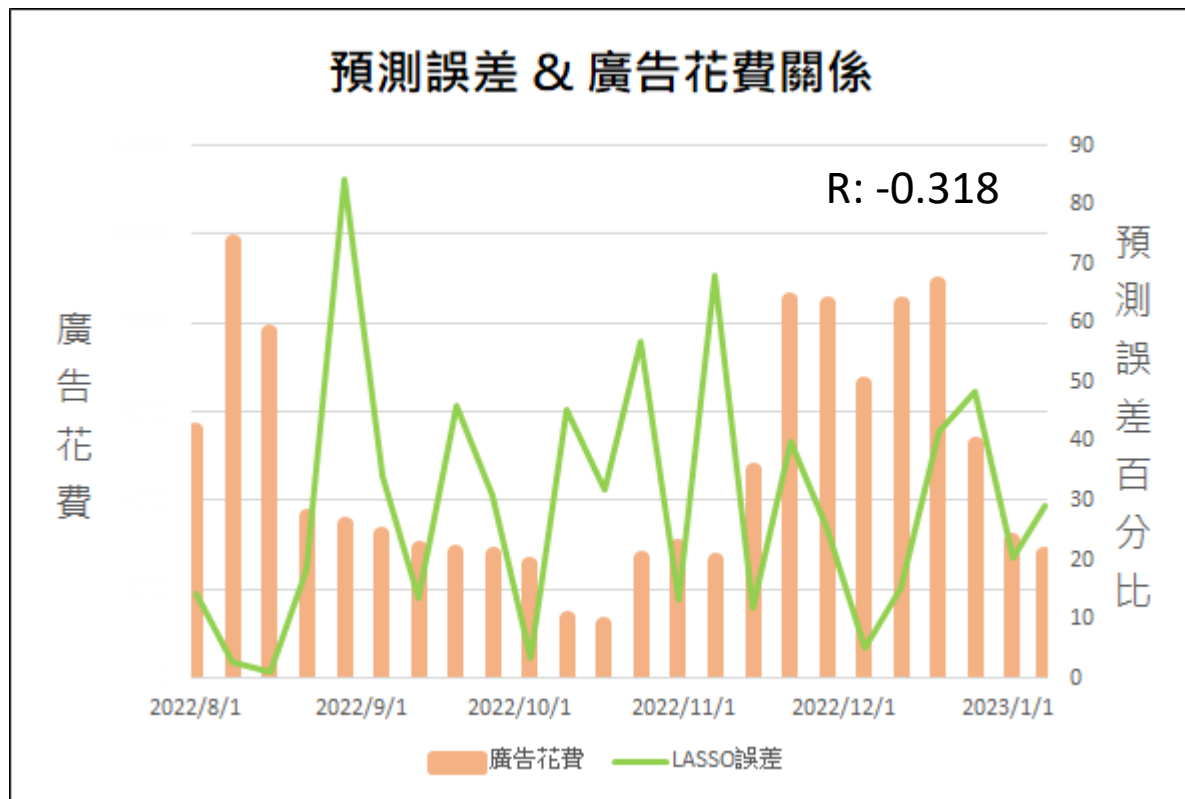
每週30%的誤差還是有點大

可能誤差大的原因 & 數據優化

1. 是不是有投放廣告的區間誤差較大？
2. 有觀察出實際180日金額越低誤差越大的狀況
 - 誤差跟實際金額的相關性？
金額低 => 人數少
 - 是否人數越少越不穩定，觀察WAU、付費人數、廣告人數相關性
3. 人數不夠的話，分雙週看、分月看？

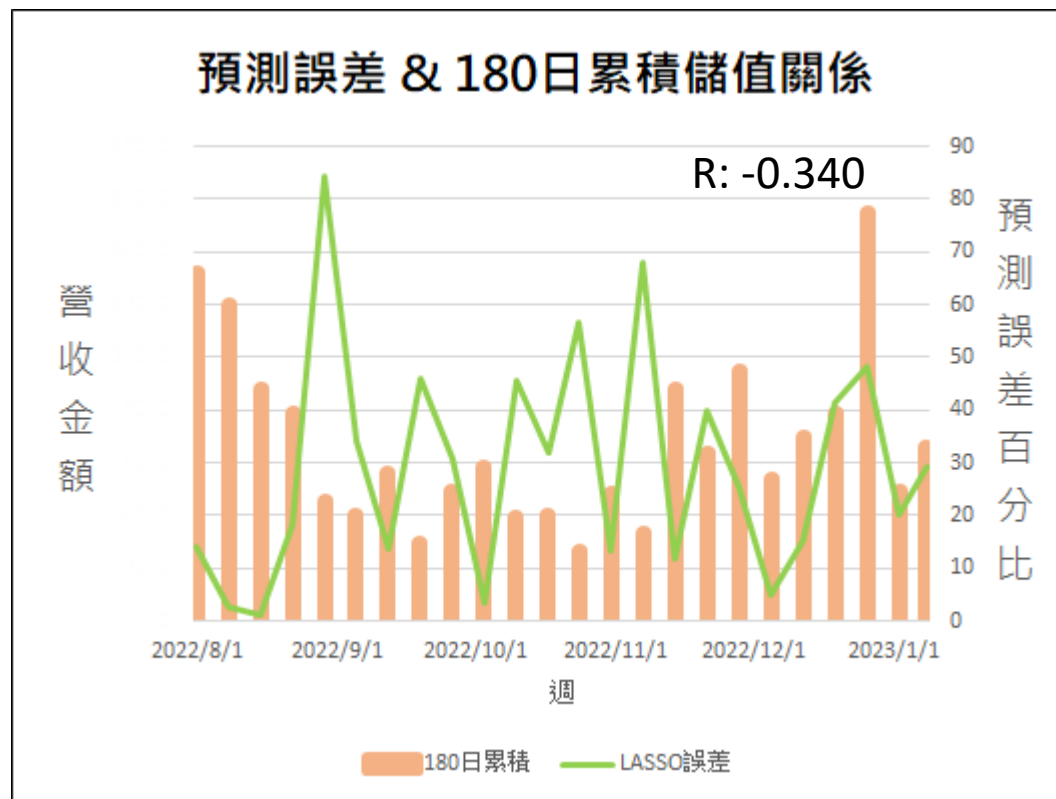


投放廣告的區間誤差(%)比較大?




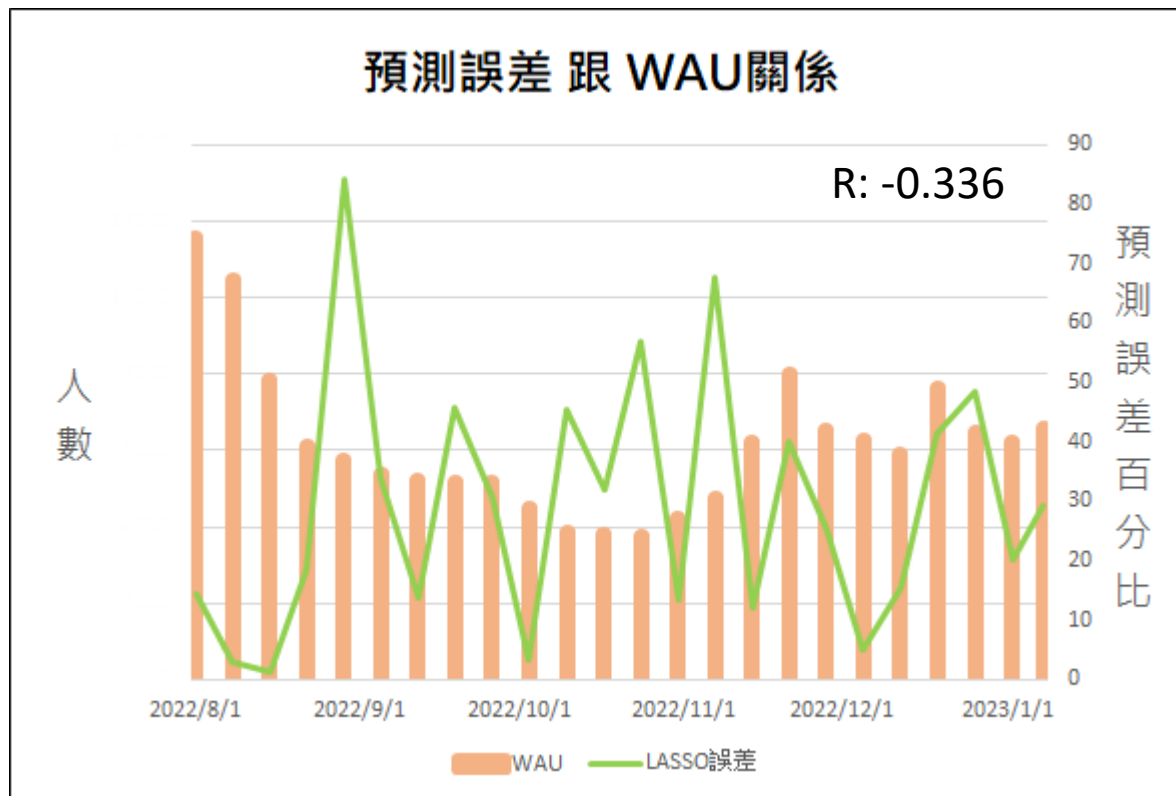
- 預測誤差跟廣告花費的相關性R為 -0.318，呈現中低度相關，表示**花費越多誤差越小**
- 跟ASA花費的相關性R為 0.049，無相關

營收越低誤差(%)越大?



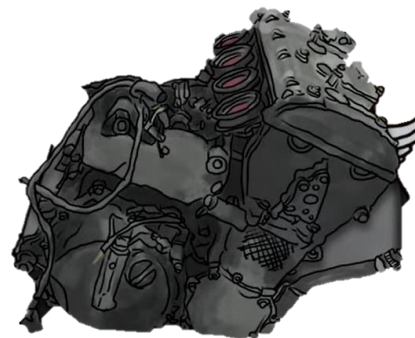
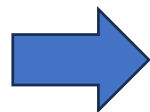
- 預測誤差跟營收的相關性R為 -0.340，呈現中低度相關，表示**營收越高誤差越小**

人數越少誤差(%)越大? 



- 預測誤差跟WAU的相關性R為 -0.336，呈現中低度相關，表示當週人數越多誤差越小

營收越小誤差越大基本原因



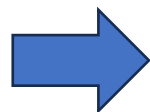
引擎大修

實際金額: 100,000

我猜金額: 90,000

誤差: 10%

絕對誤差: 10,000



引擎大修

實際金額: 20,000

我猜金額: 10,000

誤差: 50%

絕對誤差: 10,000

- 金額越小誤差越大
- 使用絕對誤差來輔助評估

誤差 & 各指標相關性

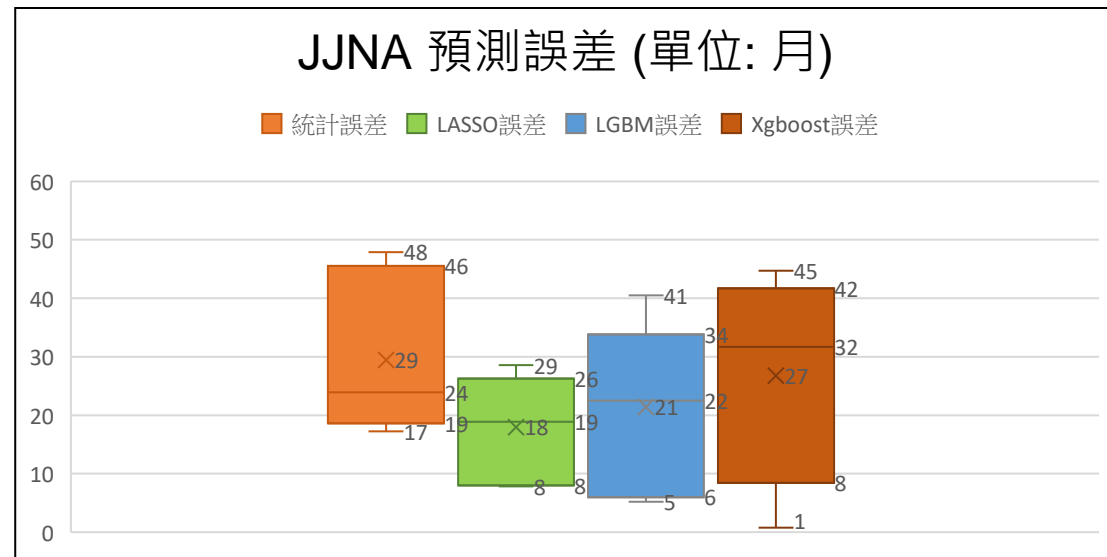
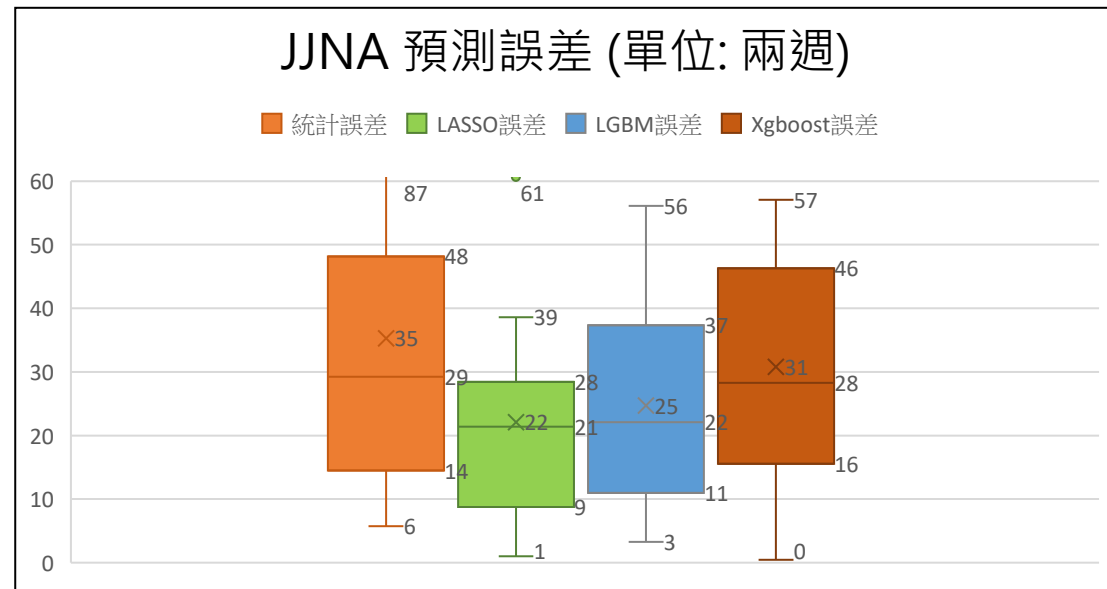
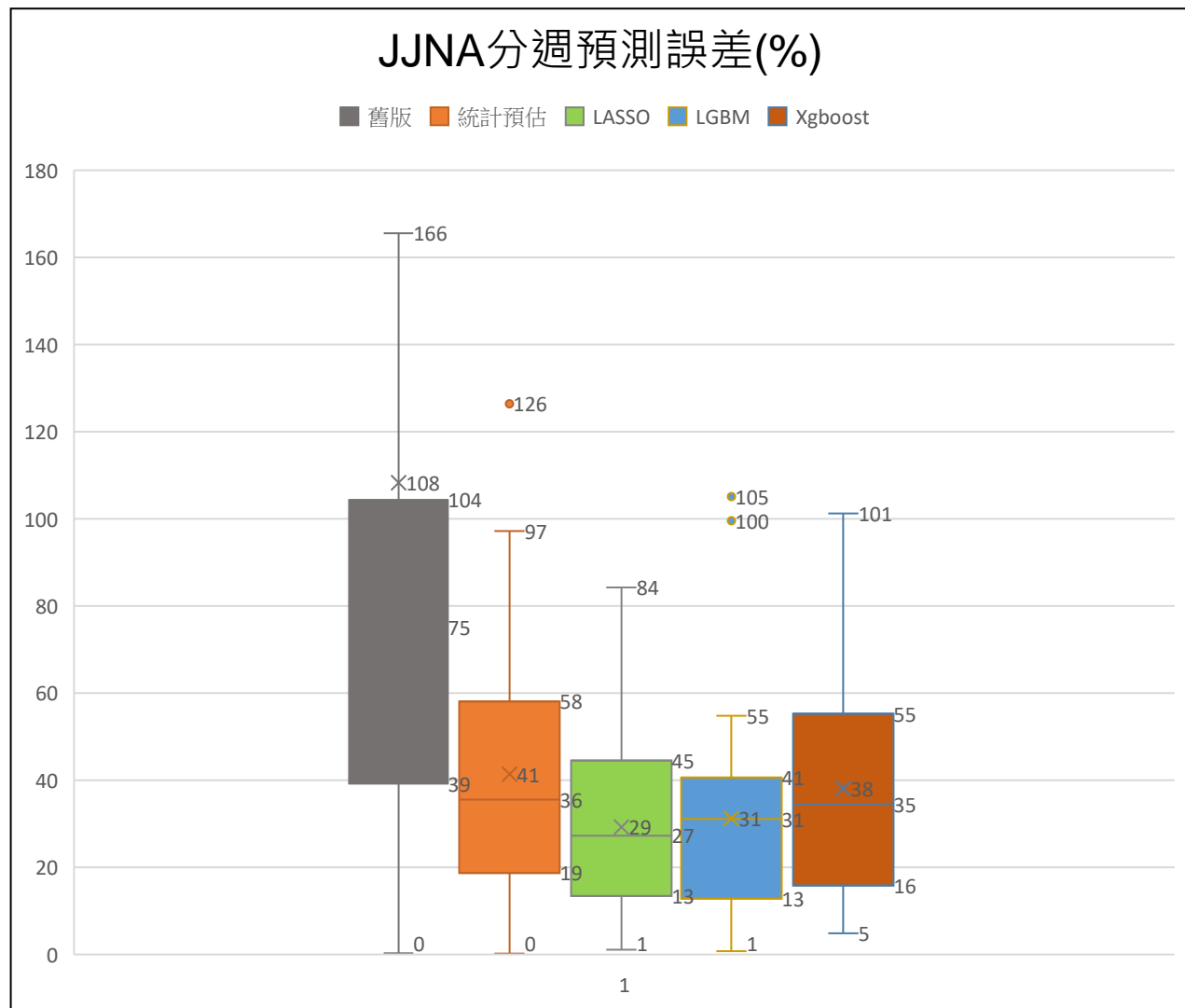
	誤差(%) 相關性	絕對誤差 相關性
總金額	-0.34	0.41
7日付費人數	-0.28	0.19
7日廣告人數	-0.35	-0.06
WAU	-0.34	0.06
廣告花費	-0.32	0.06
ASA廣告花費	0.05	0.15
非ASA廣告花費	-0.35	0.03

分雙週、月

- 人數越少、金額越少百分比誤差會變大
- 金額越高絕對誤差會變大，也能得知人數多寡不影響絕對誤差



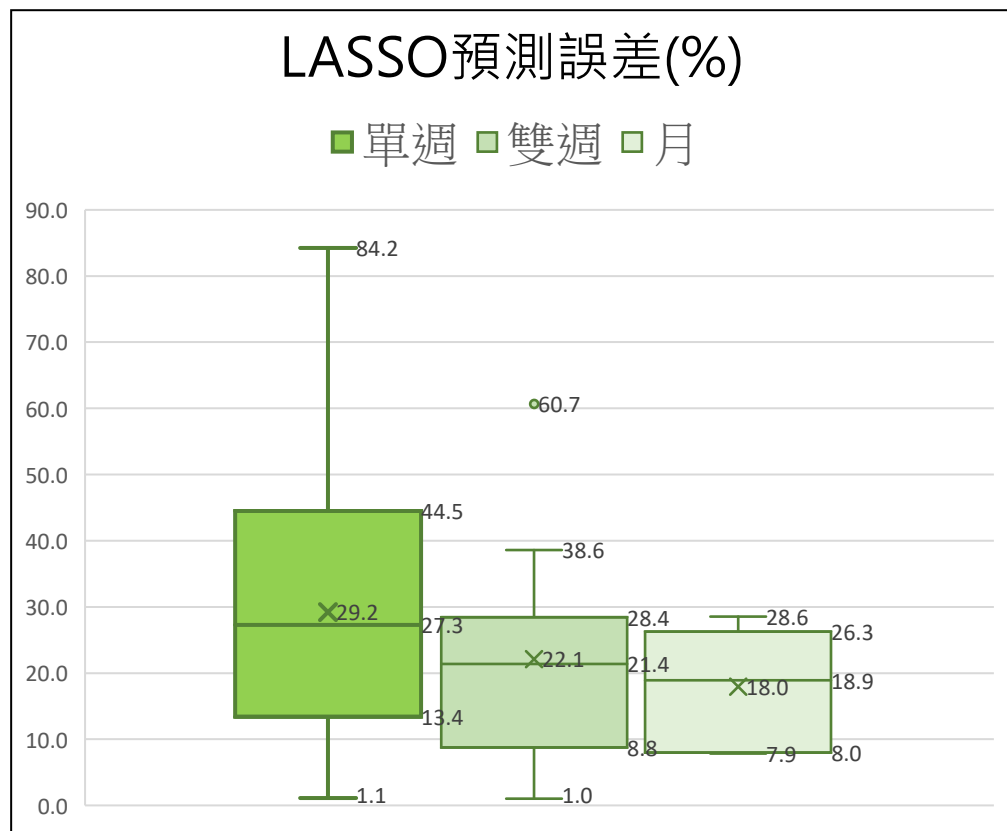
分雙週跟分月結果是否較穩定?



LASSO預測誤差-分平台

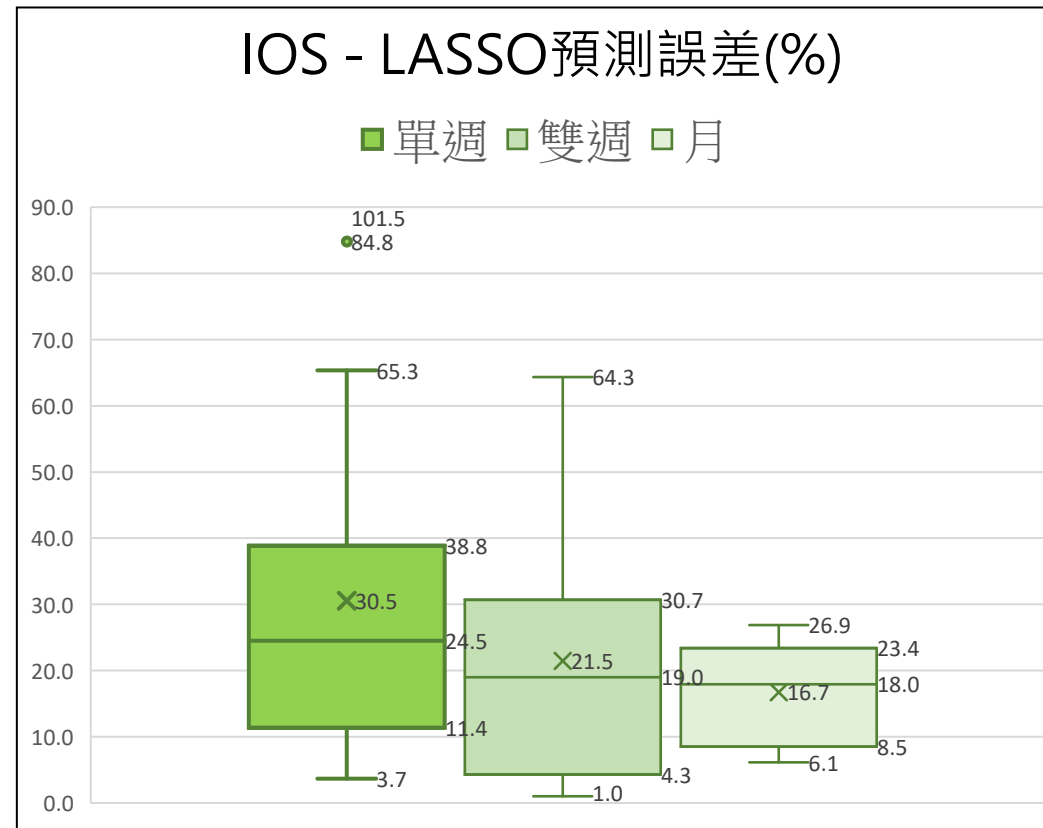
訓練模型: 全平台

預測: 全平台



- 人數越多 > 誤差(%)越穩定

預測: IOS



- IOS預測結果跟全平台預測結果差不多

單週評估

R: 0.64

分週	人數	實際180日營收	預測營收	誤差(%)	絕對誤差
2022/8/1				13	
2022/8/8				-13	
2022/8/15				-13	
2022/8/22				9	
2022/8/29				102	
2022/9/5				31	
2022/9/12				6	
2022/9/19				37	
2022/9/26				25	
2022/10/3				4	
2022/10/10				65	
2022/10/17				34	
2022/10/24				54	
2022/10/31				11	
2022/11/7				85	
2022/11/14				14	
2022/11/21				32	
2022/11/28				-35	
2022/12/5				-8	
2022/12/12				8	
2022/12/19				39	
2022/12/26				-54	
2023/1/2				16	
2023/1/9				24	

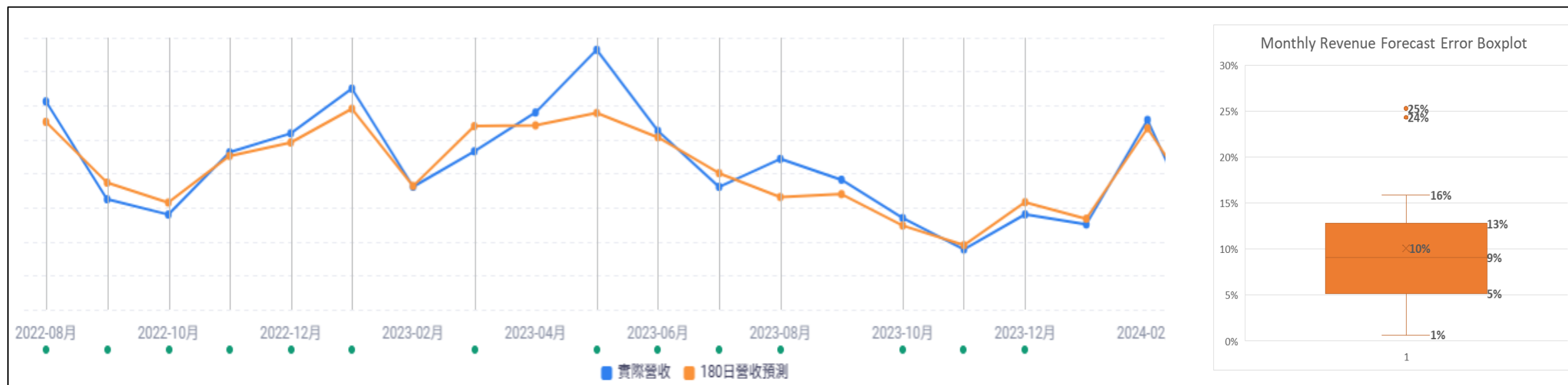
- 儲值金額越少誤差越大
- 分週誤差不穩定，人數少時誤差大
- 誤差範圍大，只有相關性高 0.6415 可以參考

雙週評估 – 投廣期間較穩定

分兩週	ASA花費	非ASA花費	人數	誤差(%)	絕對誤差
2022/8/1				-3	
2022/8/15				-2	
2022/8/29				64	
2022/9/12				17	
2022/9/26				13	
2022/10/10				48	
2022/10/24				25	
2022/11/7				33	
2022/11/21				-7	
2022/12/5				1	
2022/12/19				-23	
2023/1/2				21	

- 有投放非ASA廣告時，人數的增加可以讓誤差降低
- 反之只有ASA投放時可能誤差較大

實際使用狀況



從 2022 年 8 月到 2024 年 2 月，模型對每個月的預測誤差分布如下（以百分比表示）：

- 最小值 (Min) : 0.6%
- 第一四分位數 (Q1) : 5.9%
- 中位數 (Median) : 9.1%
- 第三四分位數 (Q3) : 12.7%
- 最大值 (Max) : 25.3%
- 預測與實際相關性 (R) : 0.933

總月份預測誤差 4.05%