



Carry4 成果說明

組員：劉宜珊 徐圓媛 廖沁旋 劉米婷



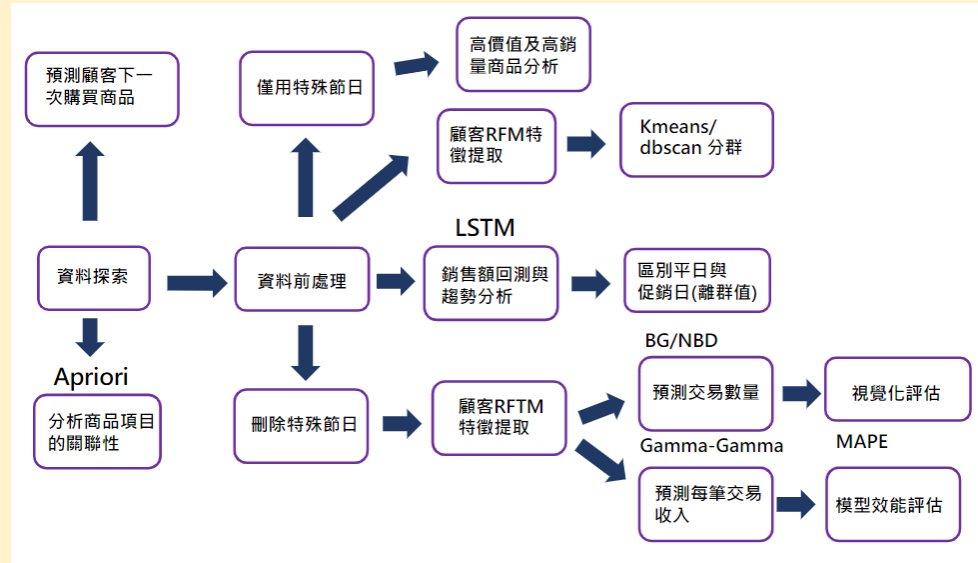
成果簡要說明

- ◆ 解題團隊：Carry4
- ◆ 出題業者：家樂福
- ◆ 業者痛點議題及期望結果簡述：
 1. 精準掌握客群特性
 2. 希望將行銷成本最大化
 3. 提升總收益

◆ 簡述使用的數據資料及技術：

1. XGBOOST客戶分群
2. LSTM 銷售額回測
3. Apriori 商品關聯性分析
4. Gamma-Gamma 交易收入預測
5. BG/NBD 交易數預測
6. 其他統計方法

解題架構圖

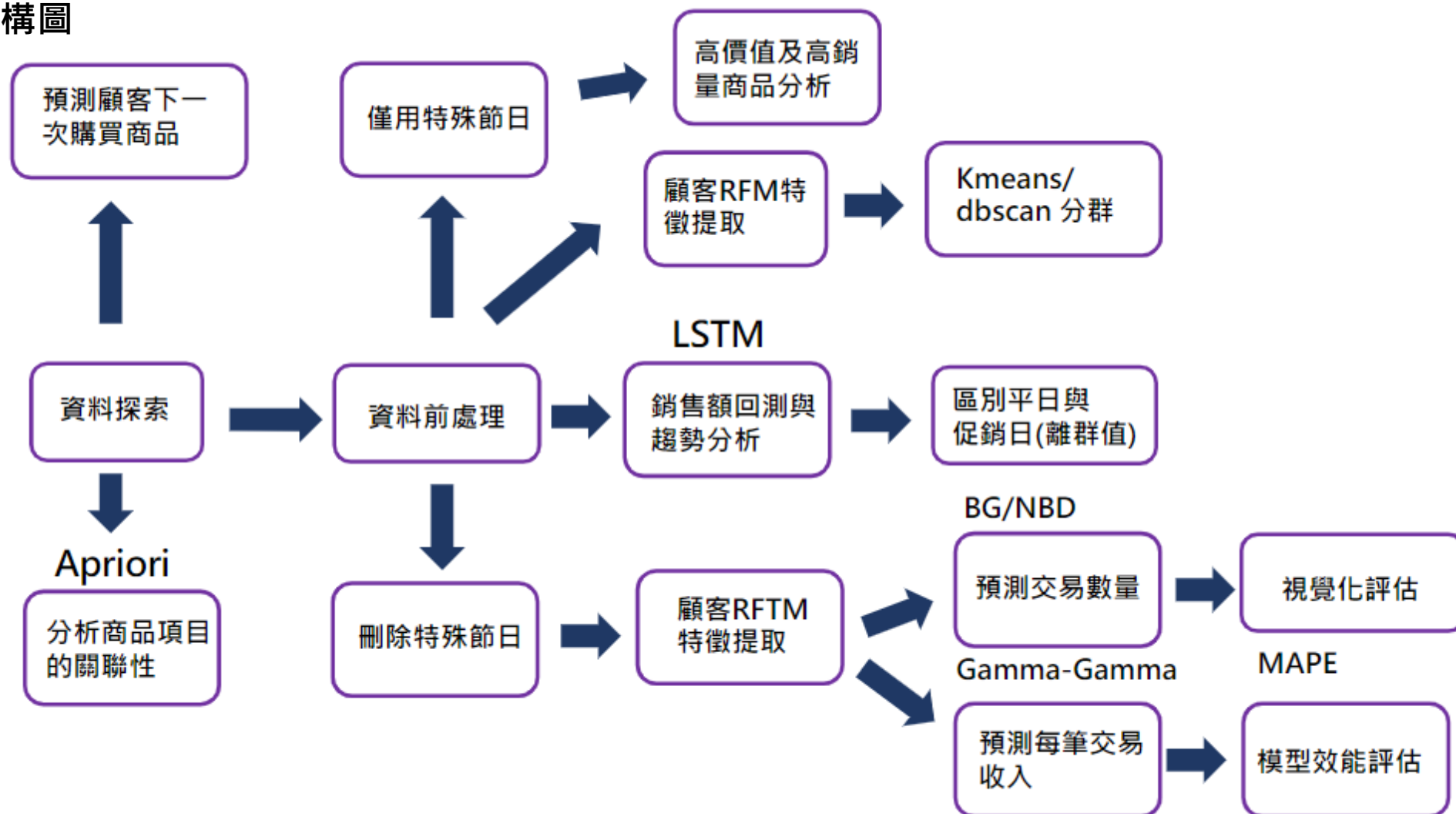


成果說明

- ◆ 解題策略：由於以IQR去除異常值的資料進行RFM顧客分群的結果並沒有無顯著差異性，我們嘗試利用機器學習的方法進行分群，對於購買分群的部分則改為探討時間及購買行為的關聯分析，並且結合統計、機器學習的技術進行交叉分析，使結果更具可信度。
- ◆ 成果亮點：考量到購買資料常以csv,xml格式儲存及處理且須保障客戶資料隱私性，此提案採用可在本地運算的Excel VBA有效統整行銷及預測資料，使資料視覺化、資訊整理與管理更加容易掌握。Excel VBA中整合此次分析的重要結果，例如促銷活動的常購商品、常態時間客戶分群清單，使抓取行銷相關的資料更為方便；針對單一客戶也能快速查詢單位時間內的預測購買次數、金額及常購商品清單，有效掌握顧客購買行為，協助檢視客戶流失狀況及客戶消費特性的轉變。



解題架構圖





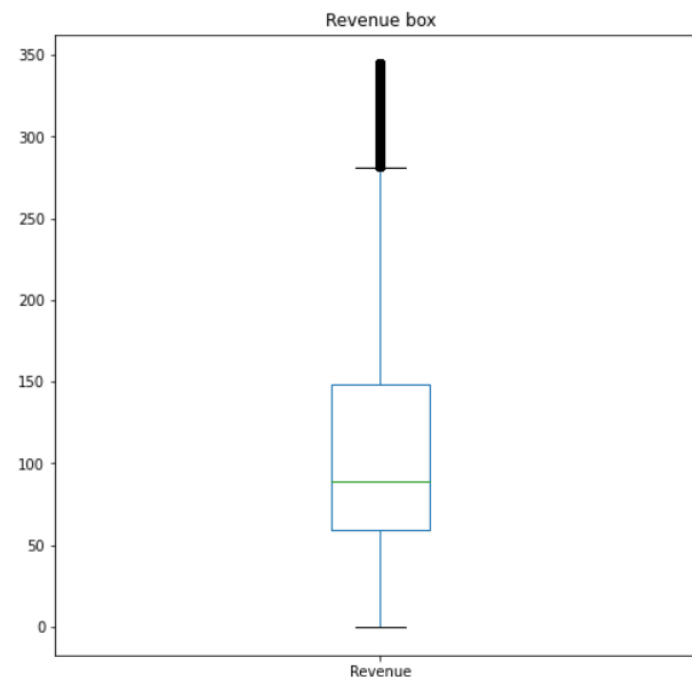
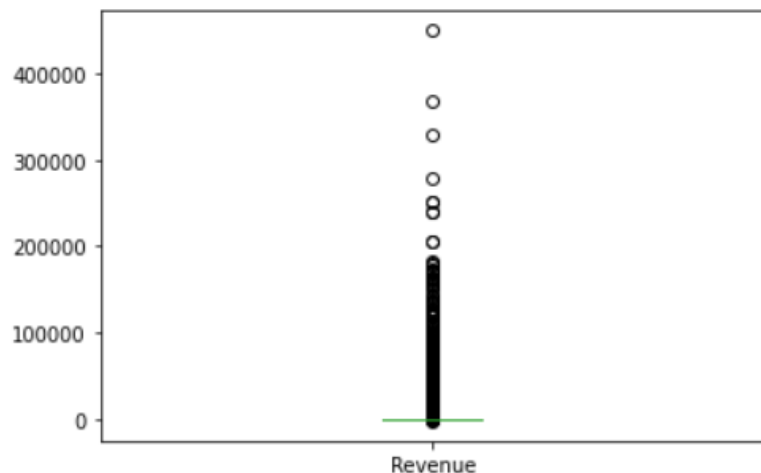
解題成果總覽

1. 顧客分群
2. 日總銷售額回測
3. 下一次購買商品預測
4. 下段時間內購買次數與金額預測
5. 特殊促銷活動日的商品行銷推薦
6. 商品項目關聯性分析
7. VBA 展示
8. 結論



1. 顧客分群

1. 將原資料以IQR去除Revenue離群值



2. 取出顧客RFM特徵

Recency: 距離最後消費日期最近

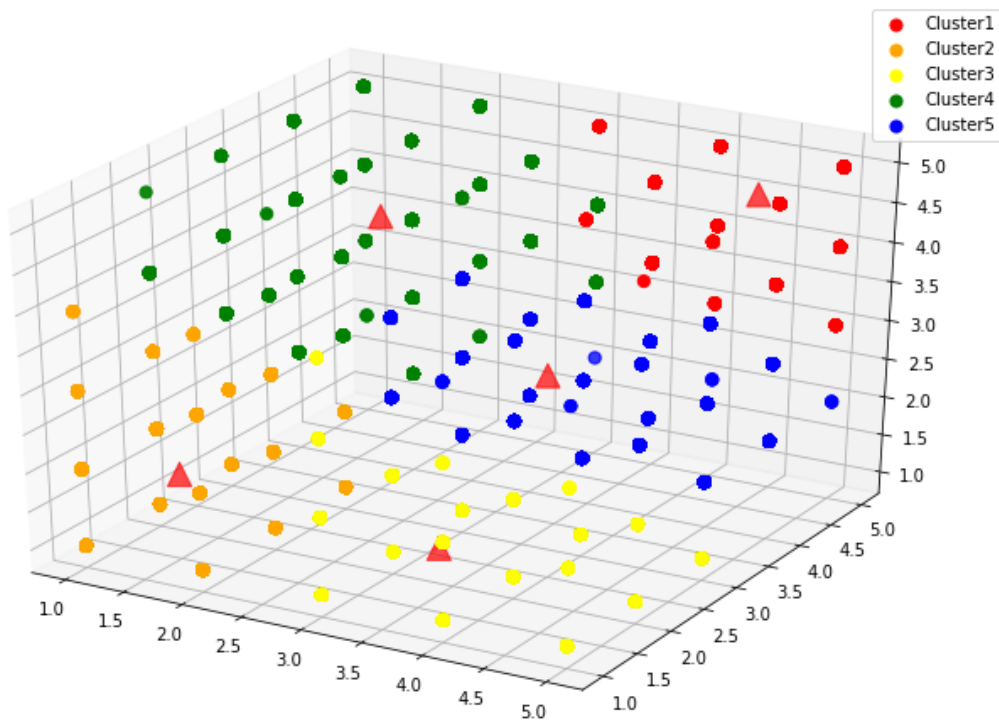
Frequency: 消費總次數

Monetary: 消費總金額

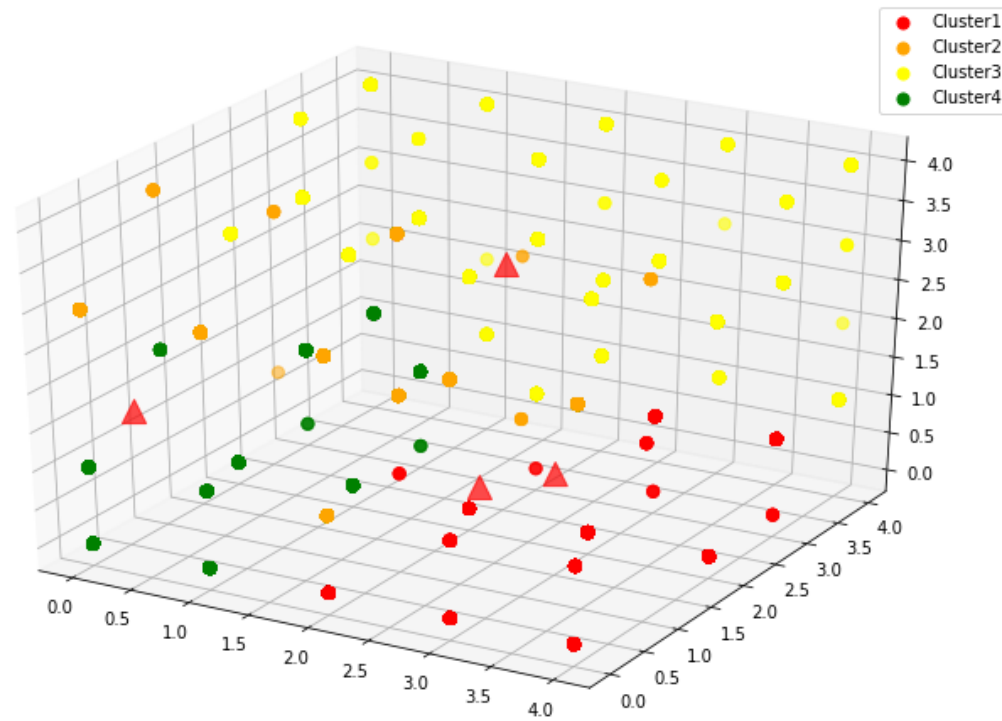


1. 顧客分群

1. Cut R、F、M by 20% of data
2. Kmeans cluster



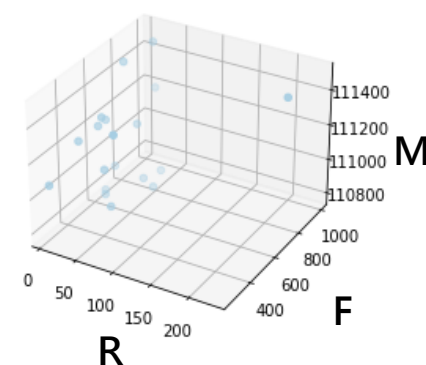
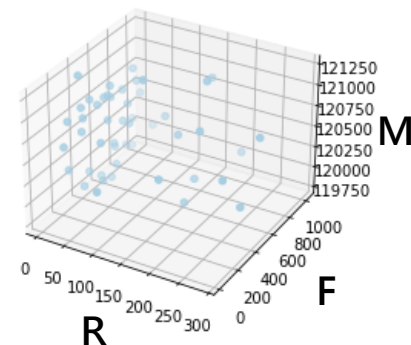
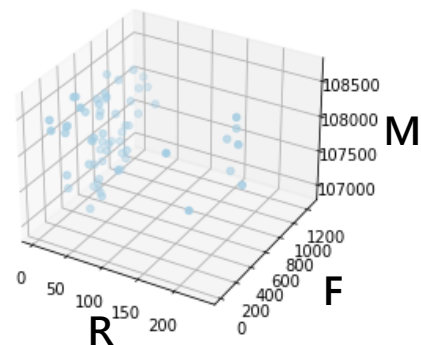
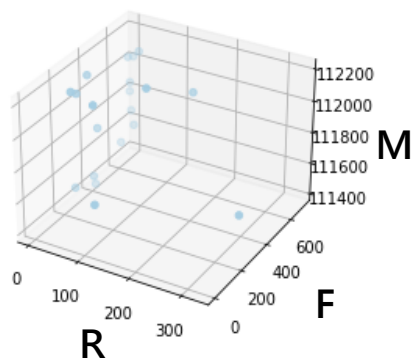
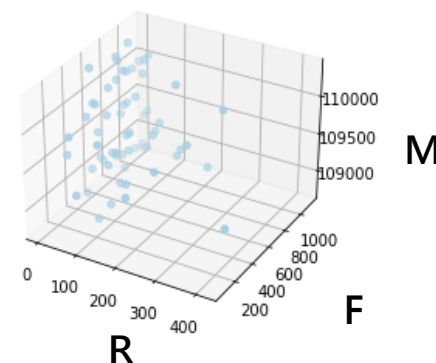
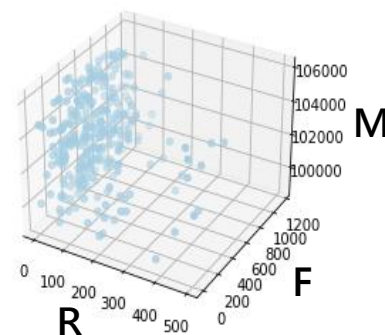
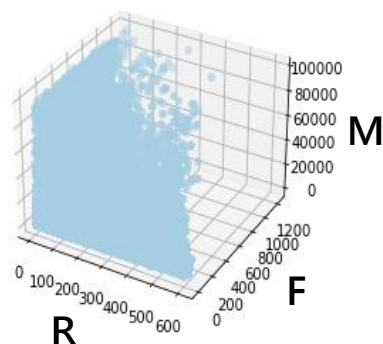
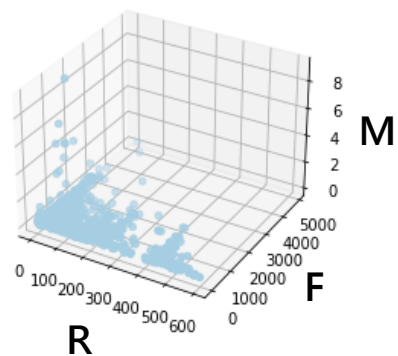
1. Kmeans cluster R、F、M
2. Kmeans cluster by previous label





1. 顧客分群

- DBSCAN: defined by dense regions
- Cluster : 8
- Noise data : 1513





1. 顧客分群

- 訓練/測試資料切割
- 顧客價值預測 以 R、F、M 加總分數計算顧客價值並預測

訓練資料：

2020/04/29~2021/08/31

測試資料：

2021/09/01~2021/12/31

SVM

	precision	recall	f1-score
High-Value	0.87	0.99	0.93
Low-Value	1.00	0.87	0.93
Mid-Value	0.86	0.96	0.91
accuracy			0.92
macro avg	0.91	0.94	0.92
weighted avg	0.93	0.92	0.92

XGBoost

	precision	recall	f1-score
High-Value	0.87	1.00	0.93
Low-Value	1.00	0.88	0.94
Mid-Value	0.87	0.96	0.91
accuracy			0.93
macro avg	0.91	0.95	0.93
weighted avg	0.93	0.93	0.93

Decision Tree

	precision	recall	f1-score
High-Value	0.90	0.76	0.82
Low-Value	0.97	0.88	0.92
Mid-Value	0.82	0.95	0.88
accuracy			0.89
macro avg	0.90	0.86	0.87
weighted avg	0.90	0.89	0.89

Logistic Regression

	precision	recall	f1-score
High-Value	0.80	0.97	0.88
Low-Value	0.99	0.76	0.86
Mid-Value	0.76	0.92	0.83
accuracy			0.85
macro avg	0.85	0.89	0.86
weighted avg	0.87	0.85	0.85

[備註]

Low value 低價值：0~1

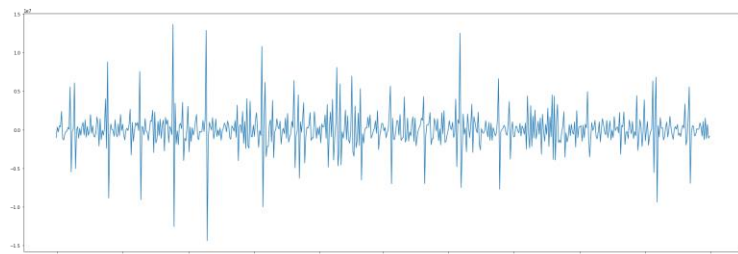
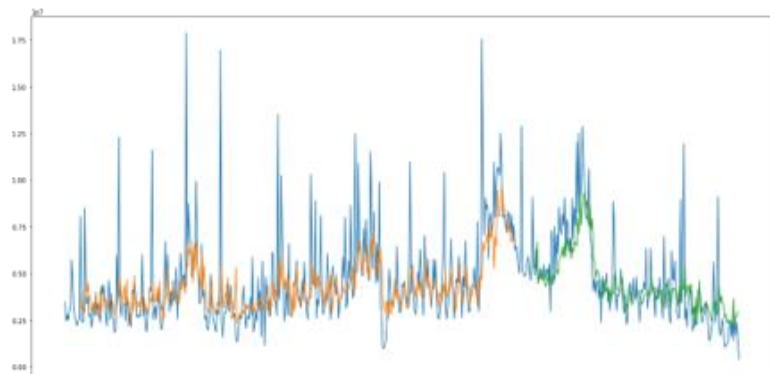
Mid value 中價值：2~4

High value 高價值：5~6



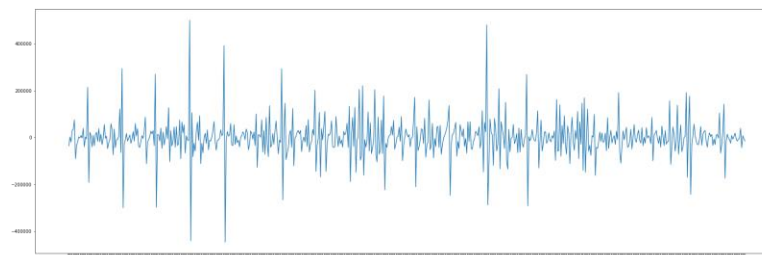
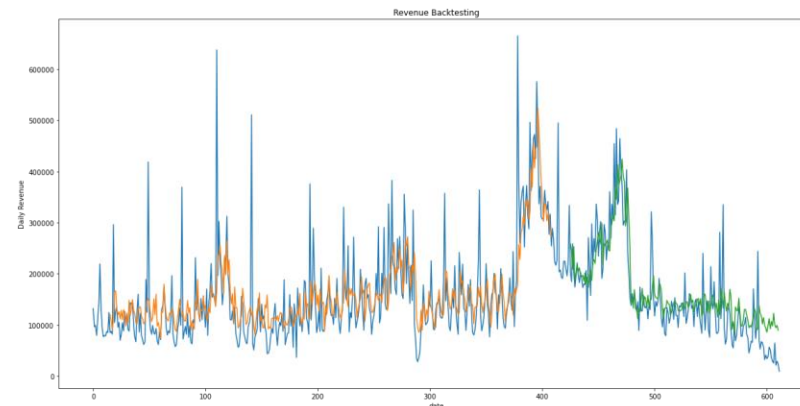
2. 日總銷售額回測

原始資料進行回測



Train Score: 2093222.33 RMSE
Test Score: 1625736.45 RMSE

去除IQR以後進行回測



Train Score: 67620.61 RMSE
Test Score: 54675.66 RMSE

去掉IQR後的回測RMSE大幅下降，意即誤差值降低，因此後續將離群值單獨操作作為銷售推薦。



3. 下一次購買商品預測

- Gensim：實現Item2vec的訓練與轉換

[概念]：

將 item 視為 word，而顧客行為序列視為一個集合，採用 SGNS 方法生成物品向量後，計算當前瀏覽物品與其它物品向量的相似度，從而找到相似物品

[實際應用]：

訓練資料	在2020.05 ~ 2021.8 (近 16 個月) 期間 若顧客曾購買該商品 利用 item2vec 模型並以餘弦定理算出與曾購買商品最相似之10個商品做為未來顧客可能購買的產品
測試資料	2021.8 ~ 2021.12 (近 4 個月)曾消費的顧客購買的實際商品
F1_score	0.35



3. 下一次購買商品預測

- 訓練資料

下一次購買的商品預測	計算 2021.05 ~ 2021.10 (近 6 個月) 曾消費的顧客消費次數最高的商品 , 做為顧客下一次購買的預測商品
實際下一次購買的商品	2021.11 ~ 2021.12 在近 6 個月曾消費的顧客實際最後一筆訂單商品
F1_score	0.41

- 繳交資料的訓練資料集

下一次購買的商品預測	<ul style="list-style-type: none">計算 2021.05 ~ 2021.12 (近 8 個月) 曾消費的顧客消費次數最高的商品 , 做為顧客下一次購買的預測商品針對近 8 個月未曾消費的顧客 , 下一次購買的預測商品則為近 6 個月所有顧客購買次數最多的商品
------------	--



4.下段時間內購買次數與金額預測

- 顧客特徵提取

Frequency：顧客重複購買的次數，也就是總購買次數減一（扣掉第一次購買），若只購買過一次的話，則該欄位值為 0。

Tenure：進行第一次購買成為顧客後，至今經過了多長的時間，單位可以是日、週、月等，在此以日為單位。

Recency：最後一次購買時，是年齡（上述的 T）為何的時候，也就是最後一次購買時間減掉第一次購買時間，如果該名顧客只購買過一次則該欄位值為 0。

Monetary value：顧客曾購買的商品之平均金額

- 訓練/測試資料切割

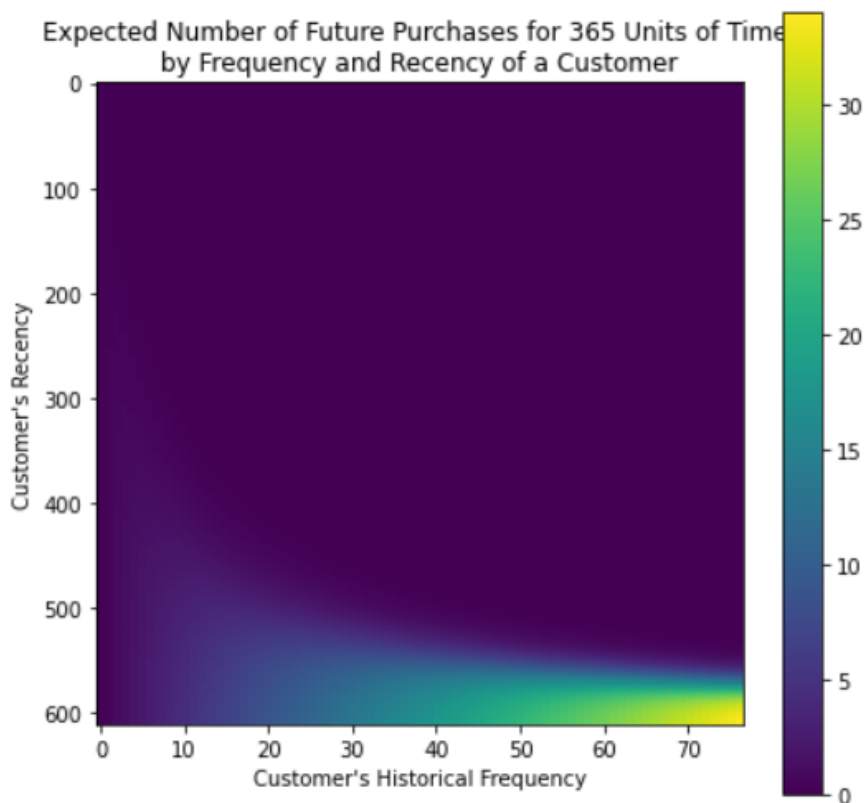
calibration period dataset (訓練資料)：2020/04/29~2021/08/31

holdout dataset (測試資料)：2021/09/01~2021/12/31

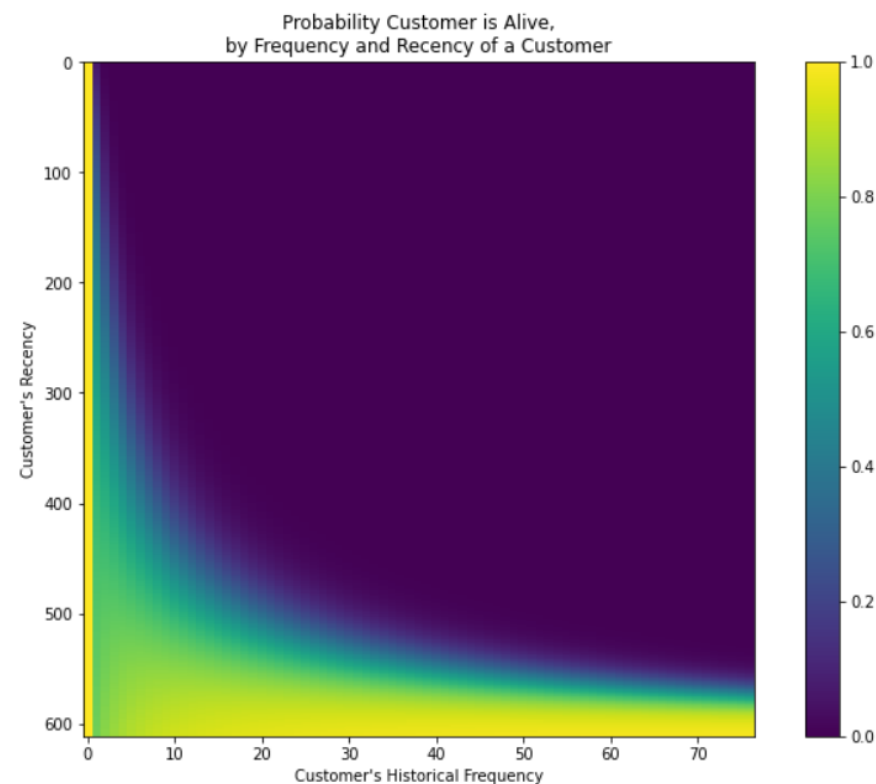


4.下段時間內購買次數與金額預測

BG/NBD Model 分析 Frequency/Recency



透過視覺化顯示未來顧客在單位時間內進行購買的次數預期值



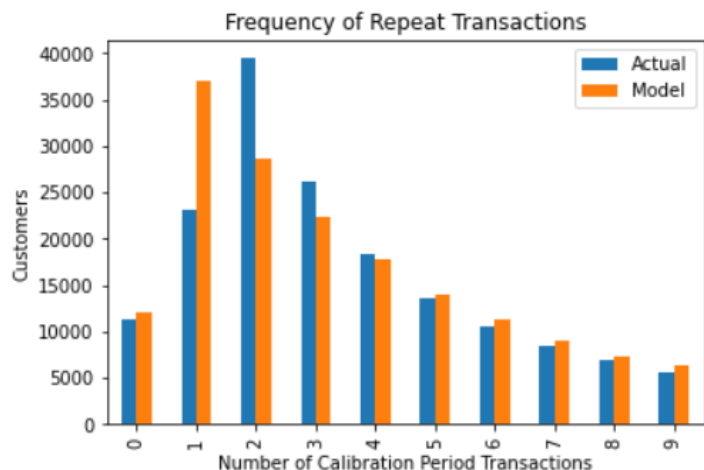
透過視覺化顯示顧客存活 (alive) 的機率



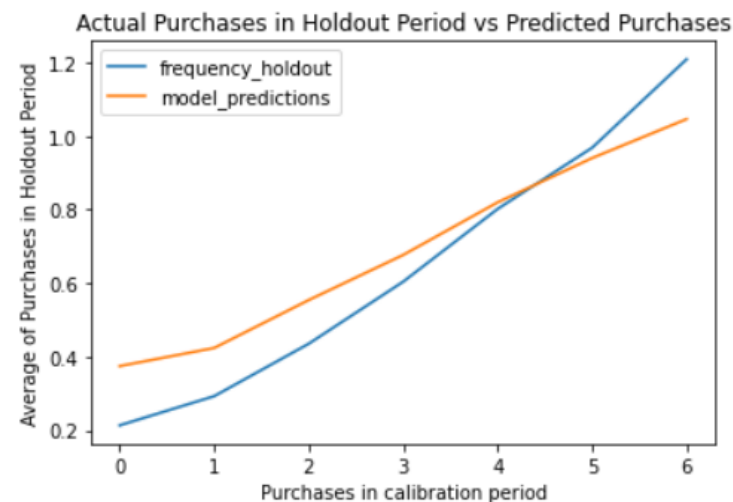
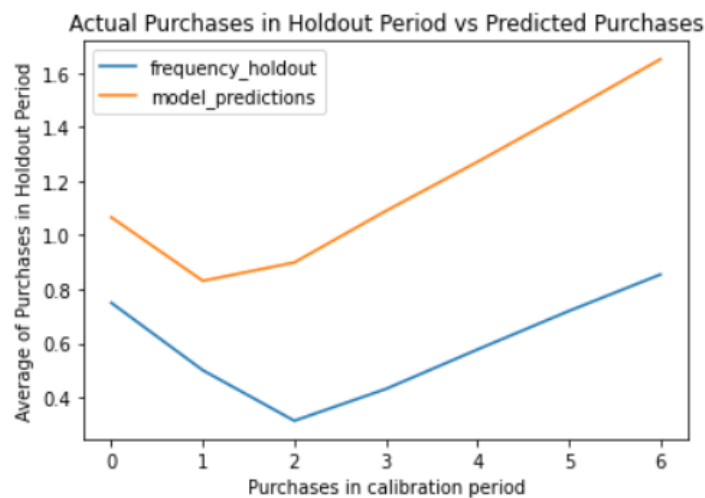
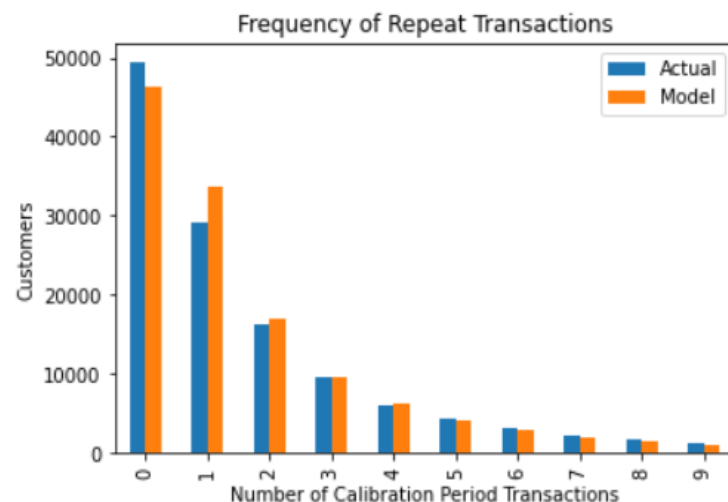
4.下段時間內購買次數與金額預測

BG / NBD Model預測購買次數驗證

原始資料



去除特殊節慶資料



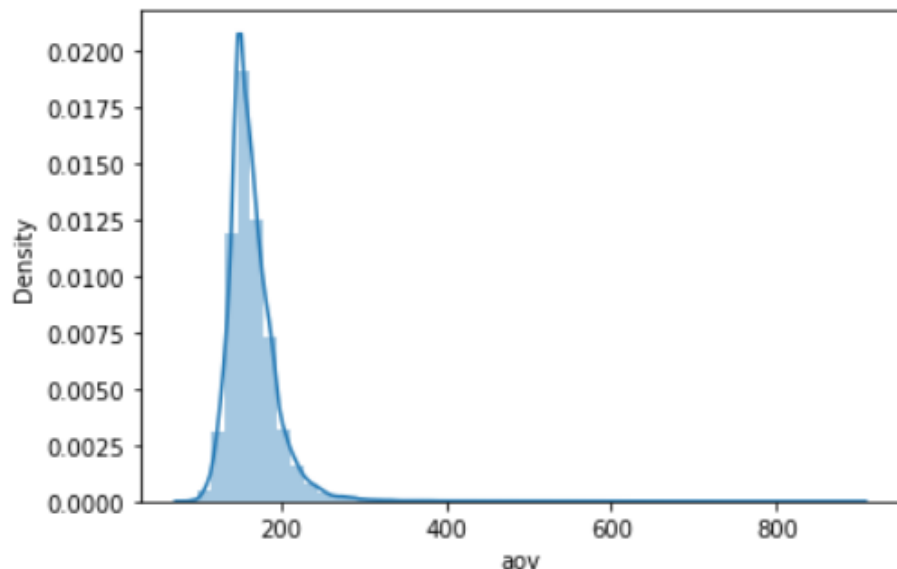
去掉特殊節慶後，購買數的預測較準確，因此在做購買數預測時應將特殊節慶促銷和平日分開做預測



4.下段時間內購買次數與金額預測

- Gamma-Gamma 預測購買平均金額

預測購買金額分布

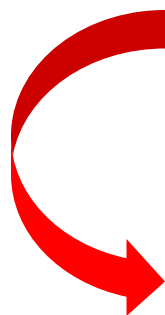
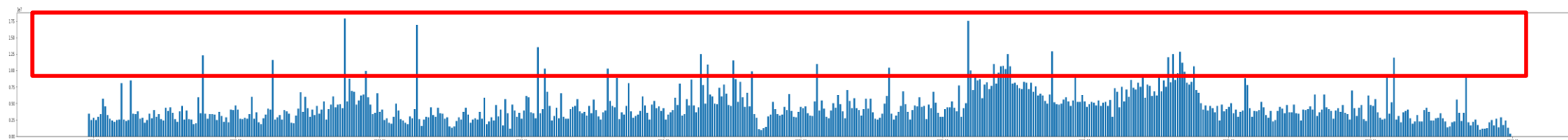


- 基於 frequency 和 monetary value 預測未來顧客可能消費的平均金額
- 以 MAPE (Mean Absolute Percentage Error) 作為預測金額的評估指標，MAPE 為 0.28

MAPE of predicted revenues: 0.28



5.特殊促銷活動日的商品行銷推薦



total_price	
order_date_notime	
2020-08-17	17899937.0
2021-05-12	17554083.0
2020-09-17	16959809.0
2020-11-08	13531202.0
2021-06-17	12905262.0
2021-08-11	12854144.0

total_price	
order_date_notime	
2021-05-29	12521499.0
2021-01-17	12475040.0
2020-06-17	12296593.0
2021-08-06	12044533.0
2021-11-11	11938691.0
2020-07-17	11599682.0

繪製每日的總銷售額，將離群值作為特殊銷售活動日，並進一步提取貢獻銷售額高的商品。



5.特殊促銷活動日的商品行銷推薦

A. 高銷售量商品列表

	product	quantity	sales_price
117	815ca534-be54-4c59-820e-d53bf094dc78	2855	65
126	7b80116d-ae90-44dc-87b5-c97f37361841	2632	76
120	7f7a7405-c43b-4cab-ac38-dc279aa12648	2180	55
195	25975452-a9ef-405b-b8c8-795f3cb01bd9	2146	75
17	ea00bae8-953f-4989-97d9-b94c39688a70	2141	12
...
59	c2e4b9e6-d0af-4420-926c-8b14971ad9b3	203	76
162	4ab8f9a3-d173-4903-a06f-ab3a5dca52f7	203	84
177	3b8dcc73-adfc-4bd5-802b-f6ea8a29858e	202	92
1	fe9f62aa-758c-4bfb-80bc-d01951399d6b	201	35
127	7a2c5cc1-c391-4edc-be03-8fc9797ffe31	201	75

231 rows × 3 columns

高銷售量商品為該月售出超過200件的商品，雖定價較低，但因銷售量多因此對於銷售額的貢獻力也大。對於行銷策略而言，適合做為組合商品、提升搭配商品的曝光度並刺激購買。



5.特殊促銷活動日的商品行銷推薦

B. 高價值商品列表

	product	quantity	sales_price
82	102339cf-b6b2-48fa-9cbb-42b6850632d3	46	6990
36	95661c46-e422-4229-bf42-04d8a548bc15	26	5956
17	c1e3053c-65bb-4afe-959a-476358c7e2f4	21	3290
32	a26883db-1eec-4312-a171-cd7ea42a6c16	61	2688
52	5dd85fa7-8bd8-4668-b25a-dce52867c612	22	2388
...
59	44632484-e64c-4533-8338-bd4f50d2f3ee	126	512
43	7c567beb-2674-4cf1-835e-4f86bec267d6	104	512
46	78f999a6-daf4-4b28-95e2-db614e0c8123	98	510
65	3b6f2a11-edf1-49c9-8080-8d2fb7614eea	20	506
80	131e47e9-5a31-431b-be05-edc8020f4ca3	81	505

89 rows × 3 columns

高價值商品為該月售出超過20件且不超過150件的商品，雖然總銷售量不高但因價格較高而有效貢獻該月的銷售哦。此商品為對於客戶而言具吸引力的商品，因此透過適合的行銷方式及配合不同的促銷活動便可以刺激購買，提高銷售總額。



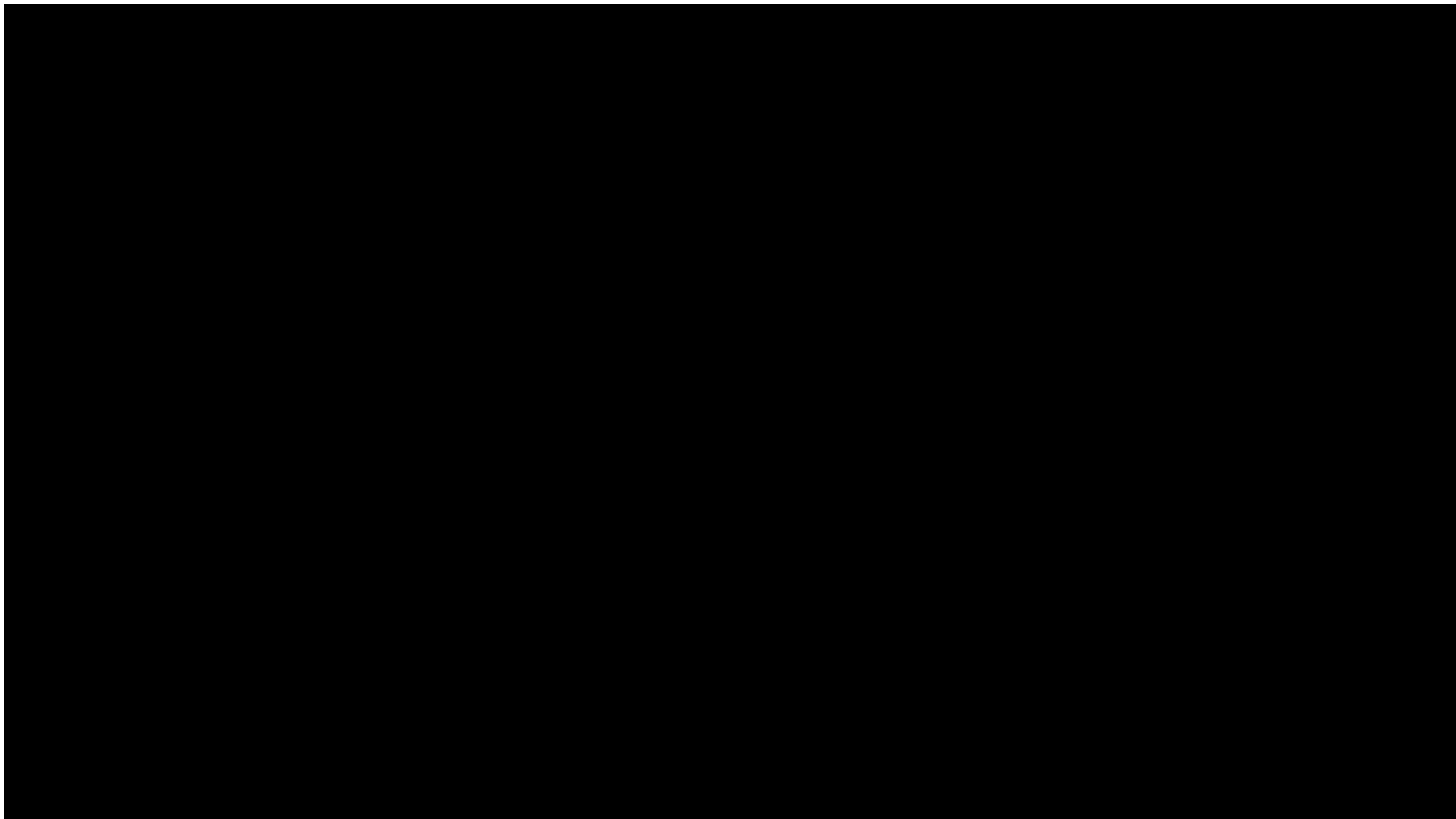
6. 商品項目關聯性分析

```
Rule: 61c5801c-15ae-4d31-8616-f8dfb9c9ad05 -> 3399e5a2-3aed-427d-97a9-5a8376a2ed46
Support: 0.03167290383822119
Confidence: 0.7551842635383339
Lift: 10.396339027048358
=====
Rule: ea00bae8-953f-4989-97d9-b94c39688a70 -> 3399e5a2-3aed-427d-97a9-5a8376a2ed46
Support: 0.03849651861460244
Confidence: 0.5299671592775042
Lift: 5.320640755398884
=====
Rule: 7b80116d-ae90-44dc-87b5-c97f37361841 -> caf67c9e-8163-4b1c-ae29-c06d40608f72
Support: 0.03043540924294156
Confidence: 0.5182364390285182
Lift: 5.443795802988527
=====
```

- 假設：
A = 61c5801c-15ae-4d31-8616-f8dfb9c9ad0
B = 3399e5a2-3aed-427d-97a9-5a8376a2ed46
- 解釋：
A對B的置信度為75%，支持度3%，提升度10。代表在購物數據中，共有3%的用戶既購買了A又買了B；同時買A的用戶有75%會購買B，A和B是正關係。
- 運用：
 - 加價購 當顧客買 a 商品時，在購物車中會出現商品 b 作為加價購且有優惠
 - 商品 b 做促銷，促使顧客在購買 a 商品時也順道購買 b



7. VBA 展示





8. 結論

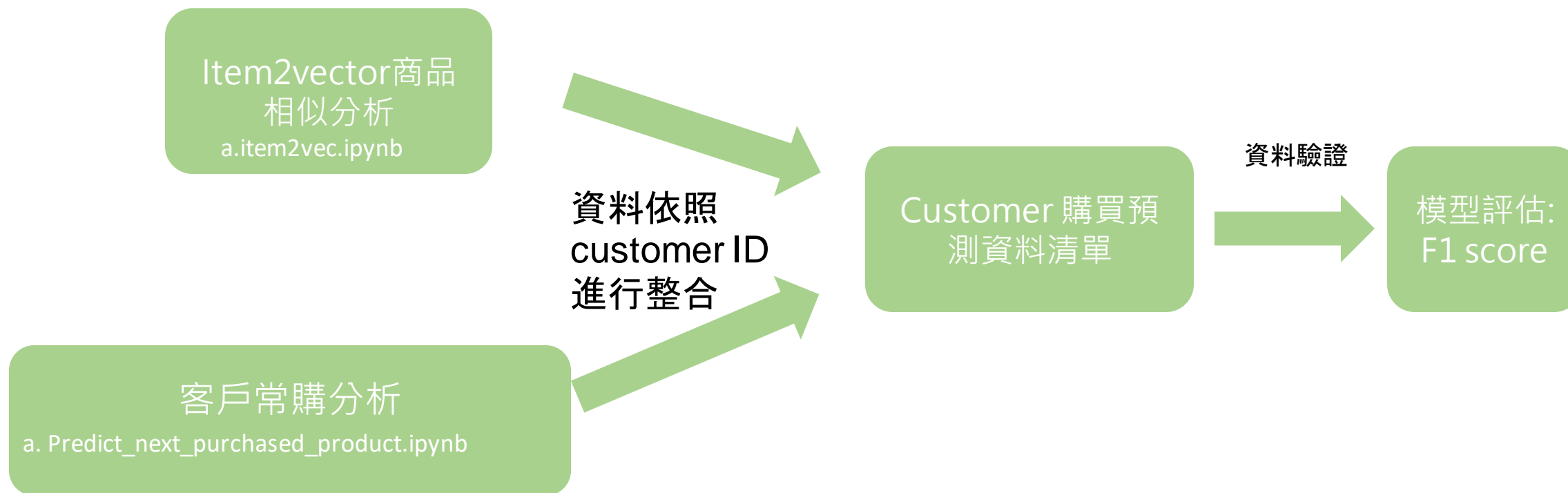
- 個別化行銷
 - ✓ Apriori 關聯分析：
當A商品被選購時可以自動推薦可能一起共買的商品
 - ✓ LSTM 趨勢分析：
區隔特殊節日及一般時間，提供不同的促銷推薦
 - ✓ 顧客分群：
分群找出高價值顧客作為行銷目標，以提高銷售額
- 未來消費預測
 - ✓ Item2vec 商品購買預測
 - ✓ 常態時間購買預測：
將客戶進行分群去預測下段時間內的購買次數及下次購買金額
 - ✓ 特殊節日購買預測：
利用統計分析找出促銷活動時貢獻銷售額高之商品
 - ✓ 下次購買時間預測：
利用統計分析找出個別顧客消費次數最高之商品

綜合結論：

常態時間時可以利用 **RTFM** 分群及 **Apriori** 關聯式分析進行精準個別行銷，而促銷活動時可以結合過去活動時間貢獻銷售額高的商品進行進一步的推薦，以有效提高各時間的銷售額。

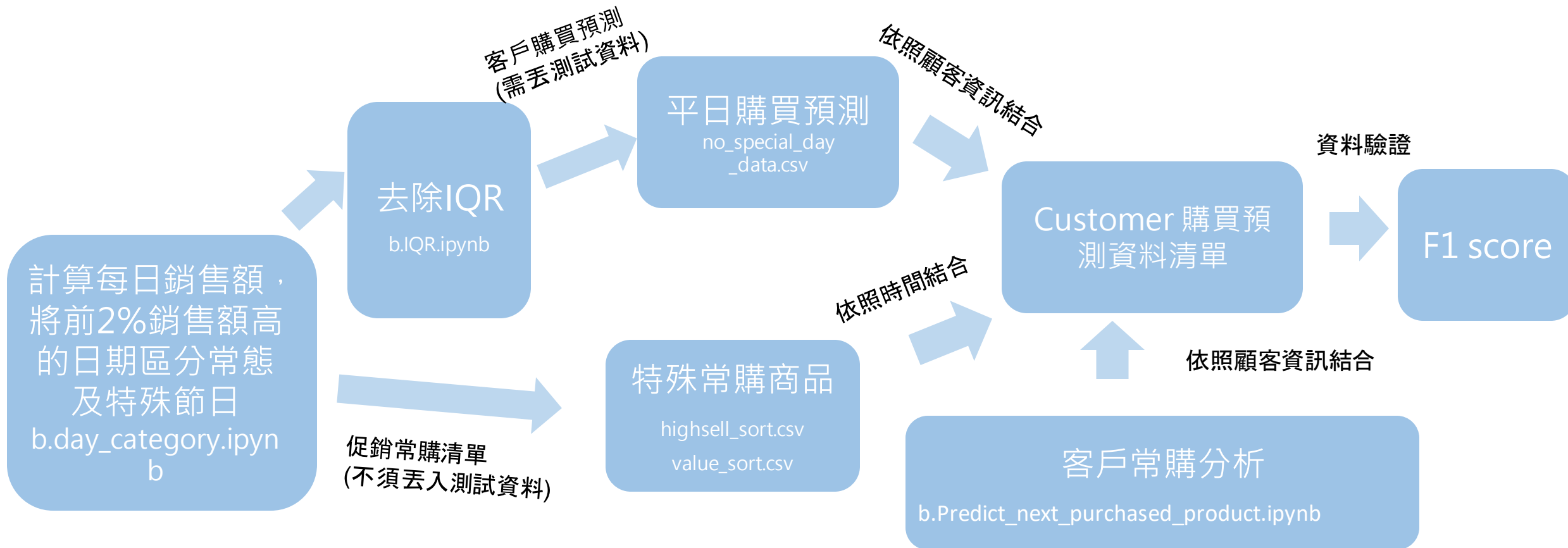


附錄a、若測試資料僅有客戶資訊



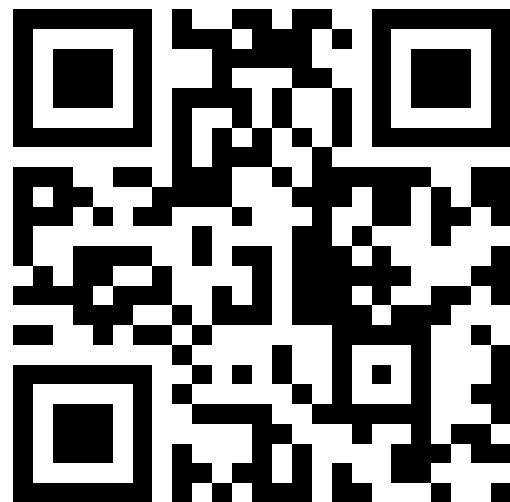


附錄b、測試資料含銷售額





附錄、Github連結



<https://reurl.cc/NRW3mk>