91APP 資料分析_大數據與商業分析
FINAL PROJECT

預測會員未來是否購買

GROUP12_徐嫵鎔 廖文豪 陳宜君 鄭皓 陳牧忠 藤田教譽

Presentation Outline

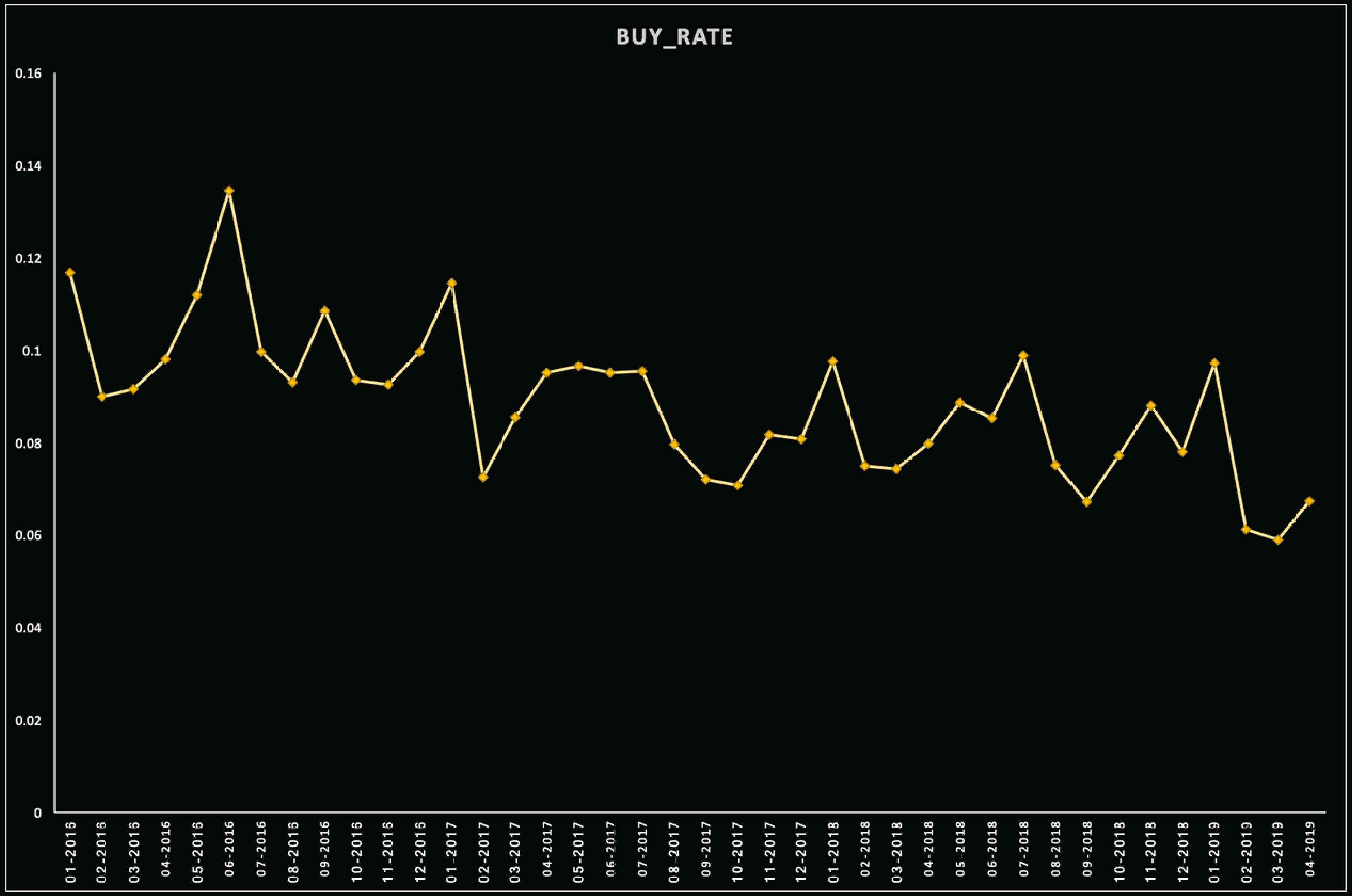
DISCUSSION POINTS

研究題目
研究目標
研究步驟
參數定義
資料分析
特徵選擇
訓練/測試場景
訓練參數
預測結果
Follow_Up

研究題目

找出關鍵少數

- 預測一個月後「購買商品」的會員商業
- 價值：減少廣告投放費用



~10%/month

研究目標

寧可錯殺一百，也不願放過一個
寧可錯殺一百，也不願放過一個
寧可錯殺一百，也不願放過一個

提高BUY RECALL RATE

研究步驟

參數定義

- Analyze 3 month data to predict 1 month
- Buy = All Order
- NoBuy= others
- Valid member
- Registered members from 2016/1/1 to the beginning of the prediction date
- At least one action
- ValidOnlineMemberID

資料分析

購買相關性

BehaviorType_Cart	0.710420
BehaviorType_Fav	0.349450
BehaviorType_Purchase	1.000000
BehaviorType_Search	0.062539
BehaviorType_ViewSalePage	0.506504
BehaviorType_ViewSalePageCategory	0.410202
TrafficSourceCategory_Direct	0.449635
TrafficSourceCategory_Email	0.112490
TrafficSourceCategory_Facebook	0.240032
TrafficSourceCategory_GoogleCpc	0.110820
TrafficSourceCategory_GoogleOrganic	0.078604
TrafficSourceCategory_Instagram	0.001484
TrafficSourceCategory_Line	0.106230
TrafficSourceCategory_LineShopping	0.044812
TrafficSourceCategory_Others	0.260757
SourceType_APP	0.450982
SourceType_WEB	0.317912
OperationSystem_Android	0.291871
OperationSystem_Chrome OS	0.004159
OperationSystem_Intel Mac OS X	0.051882
OperationSystem_Linux	0.014452
OperationSystem_Ubuntu	-0.000615
OperationSystem_Windows	0.174735
OperationSystem_iOS	0.406328

分析與購買與否之相關性

資料使用：

- 以三個月為單位，由2019/04往回切
- 整理出之BEHAVIOR FEATURE 統計數

發現：

- 客戶將商品放入購物車的次數，和購買商品的次數之相關性最為顯著。
- 作業系統中，是否使用IOS與購買次數的相關性最高。

建議：

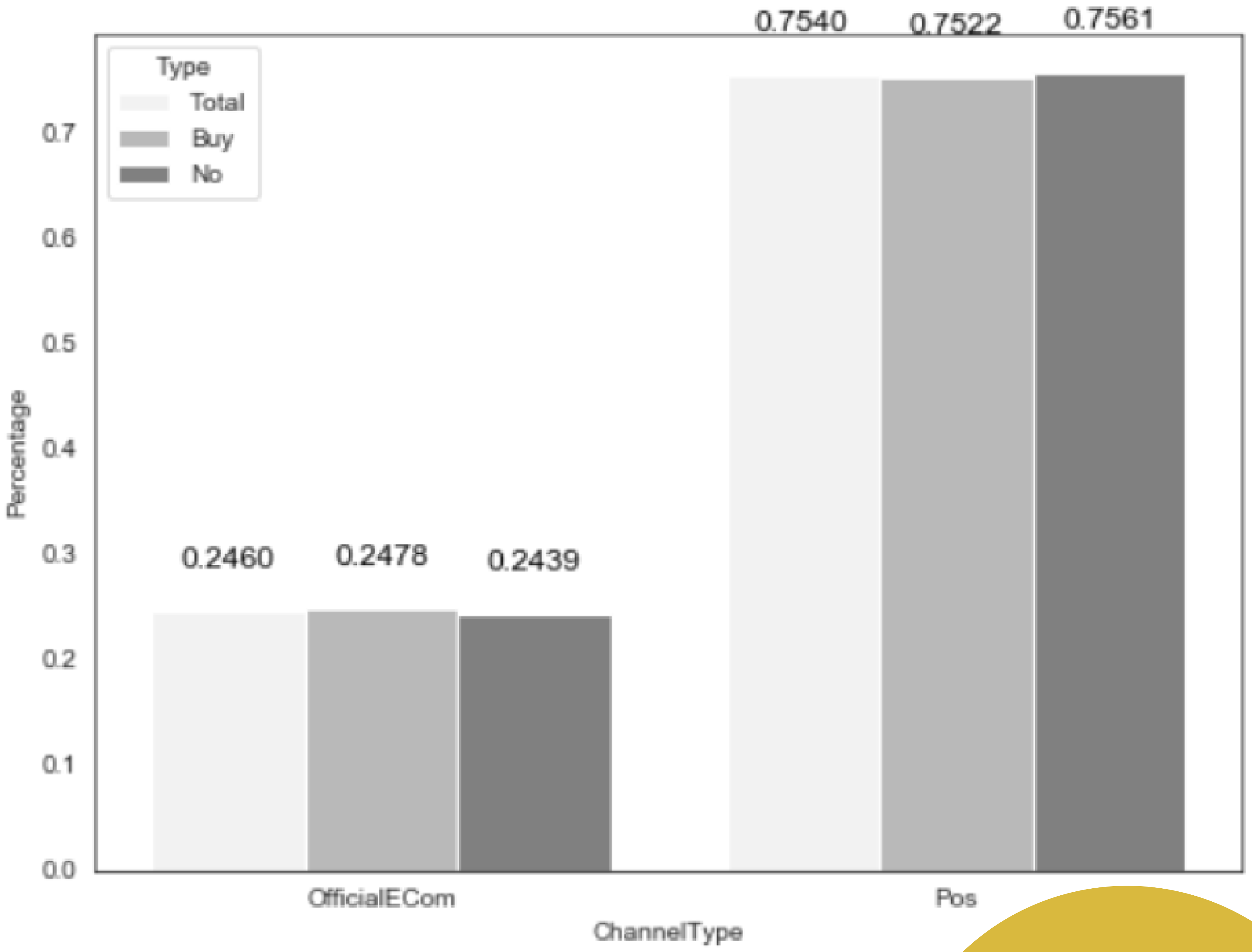
CART,VIEWSALEPAGECATEGORY,…(0.4以上的)等FEATURE與購買次數之相關性較高，在建立MODEL做預測時，優先採用這些FEATURE

資料分析

ORDER類別變數

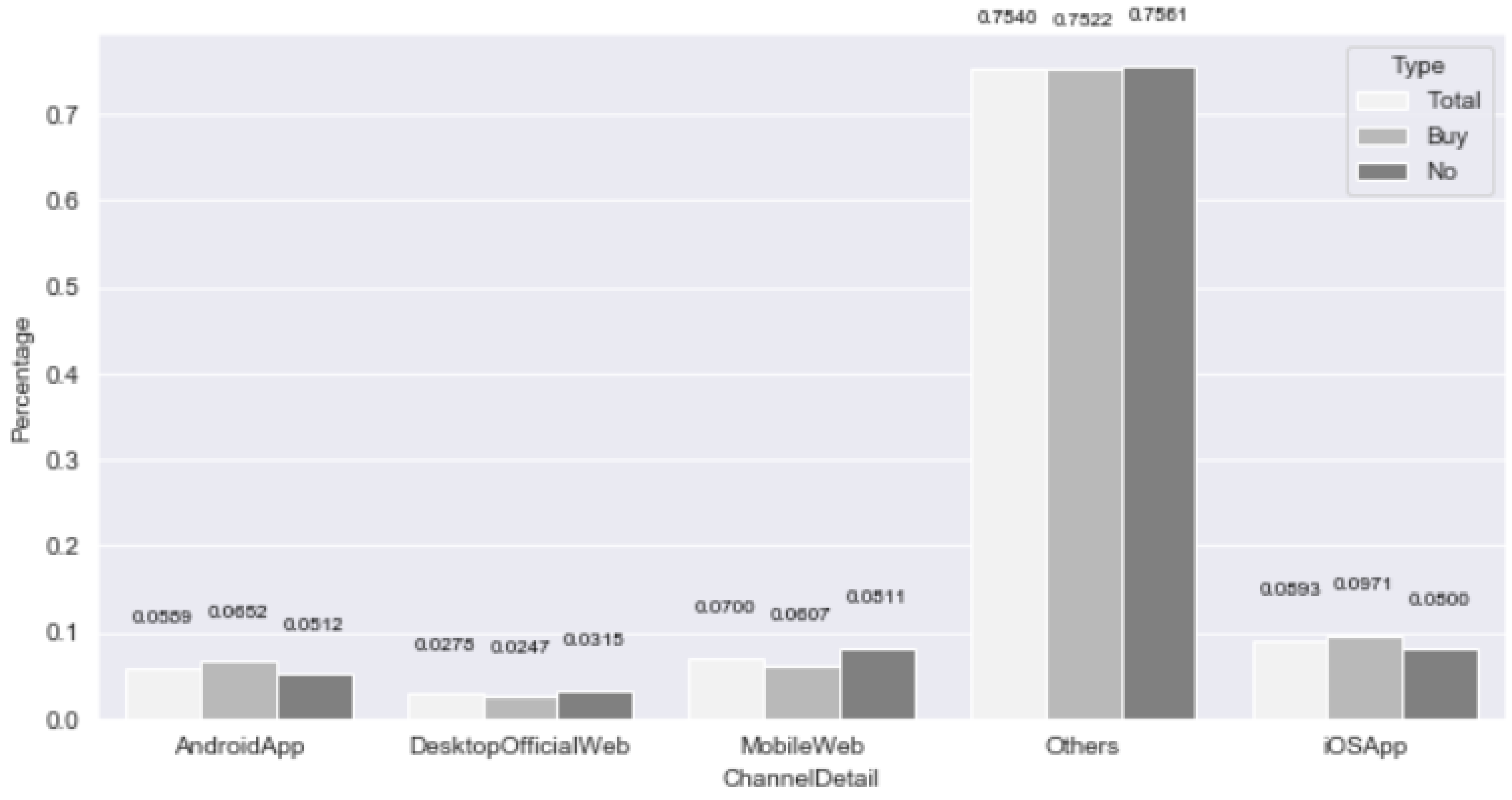
Buy：未來有購買行為
No ：未來無購買行為

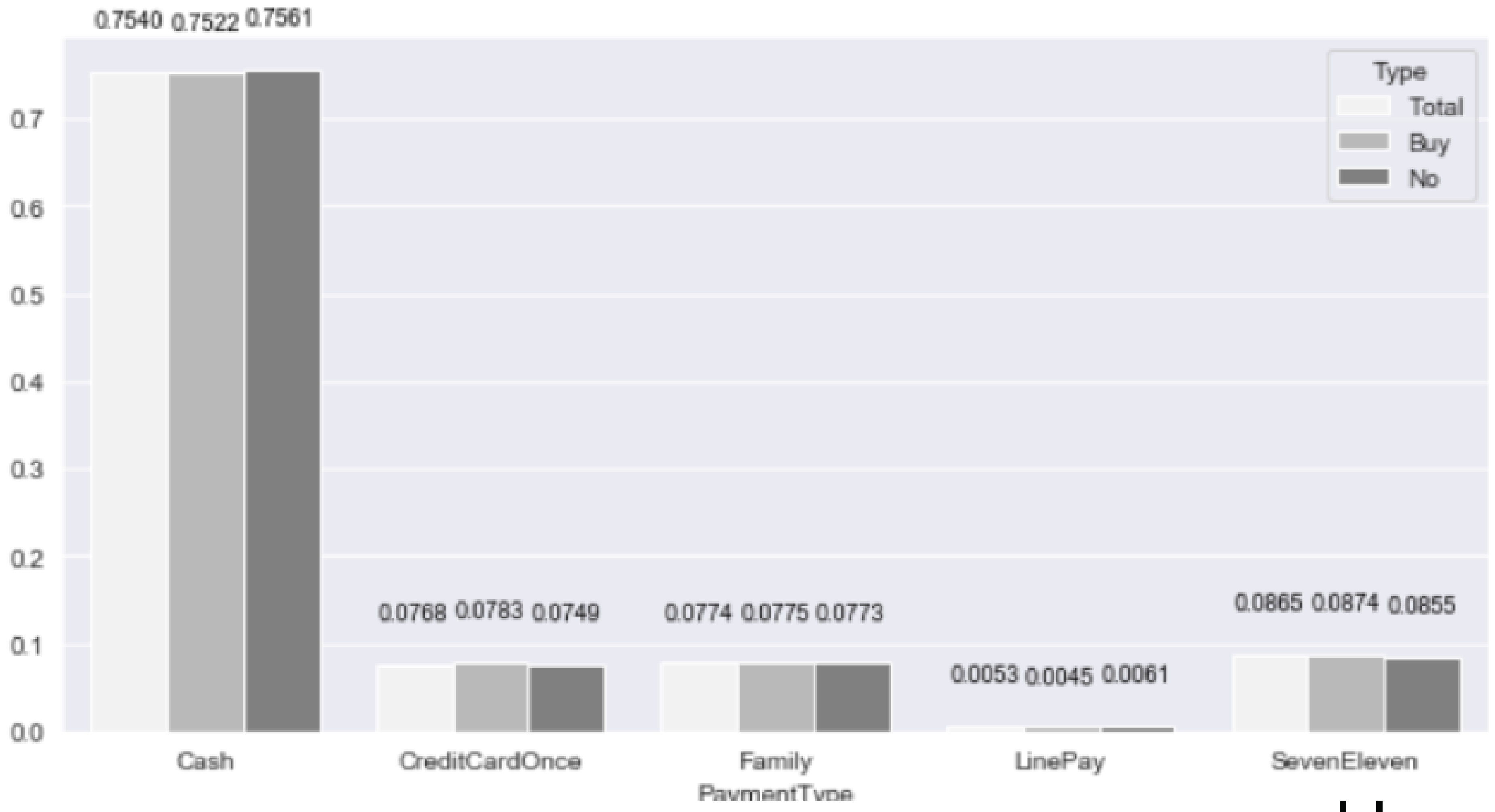
ChannelType類別差異不大



資料分析 ORDER類別變數

ChannelDetail與PaymentType同樣差異不大

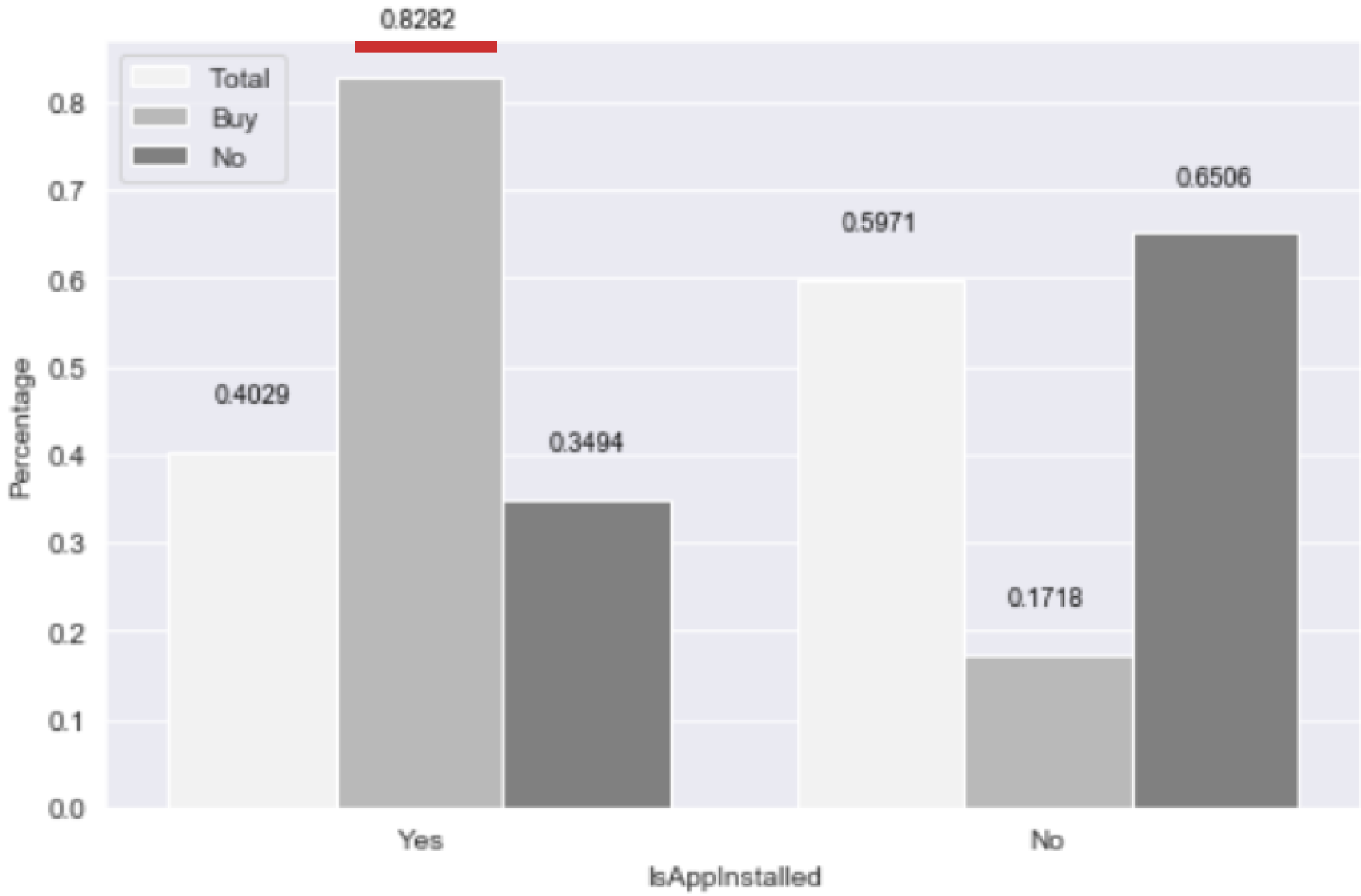




資料分析

MEMBER類別變數

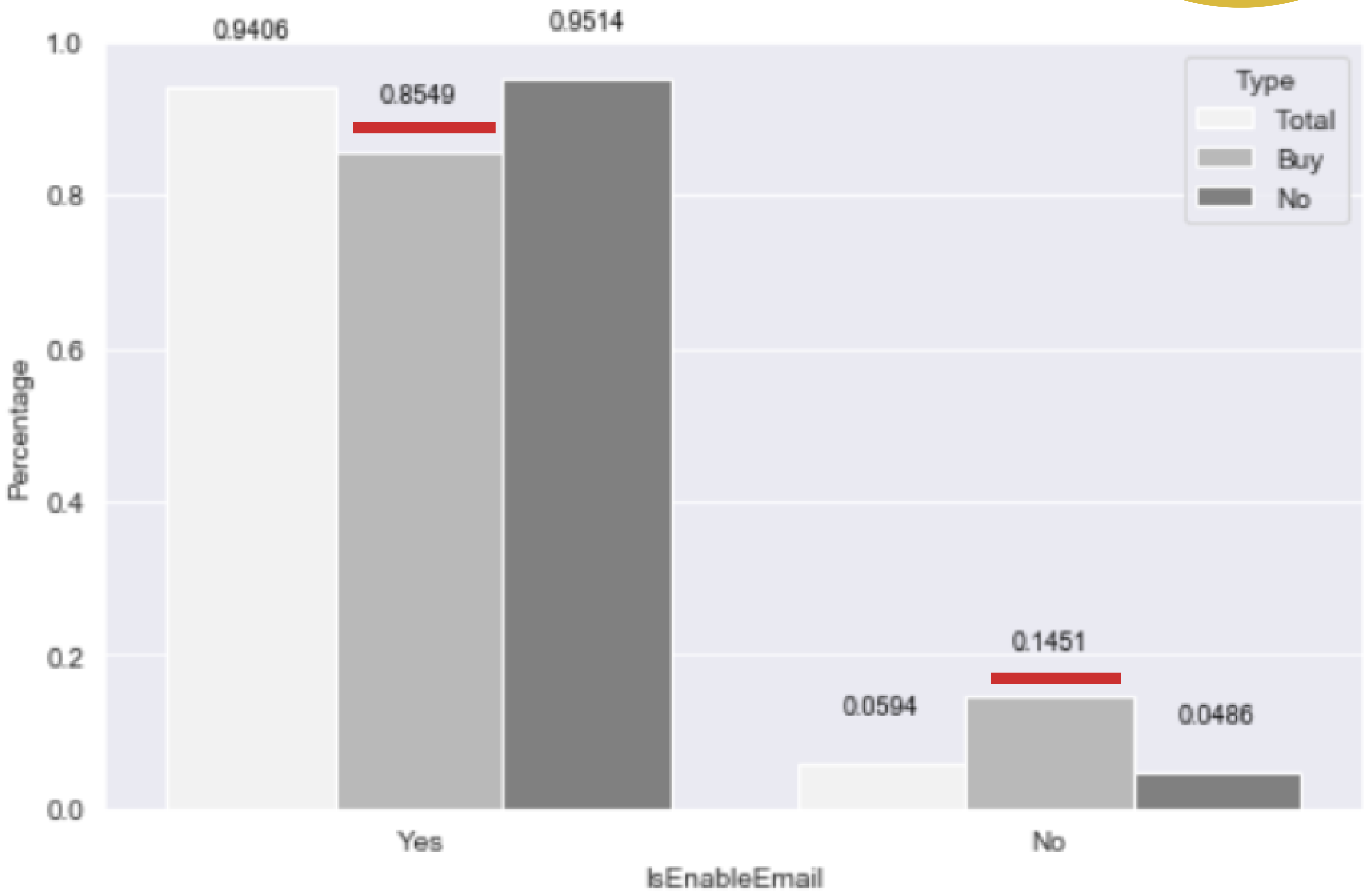
未來有購買行為的會員中，有安裝App的族群比例明顯較高



資料分析

MEMBER類別變數

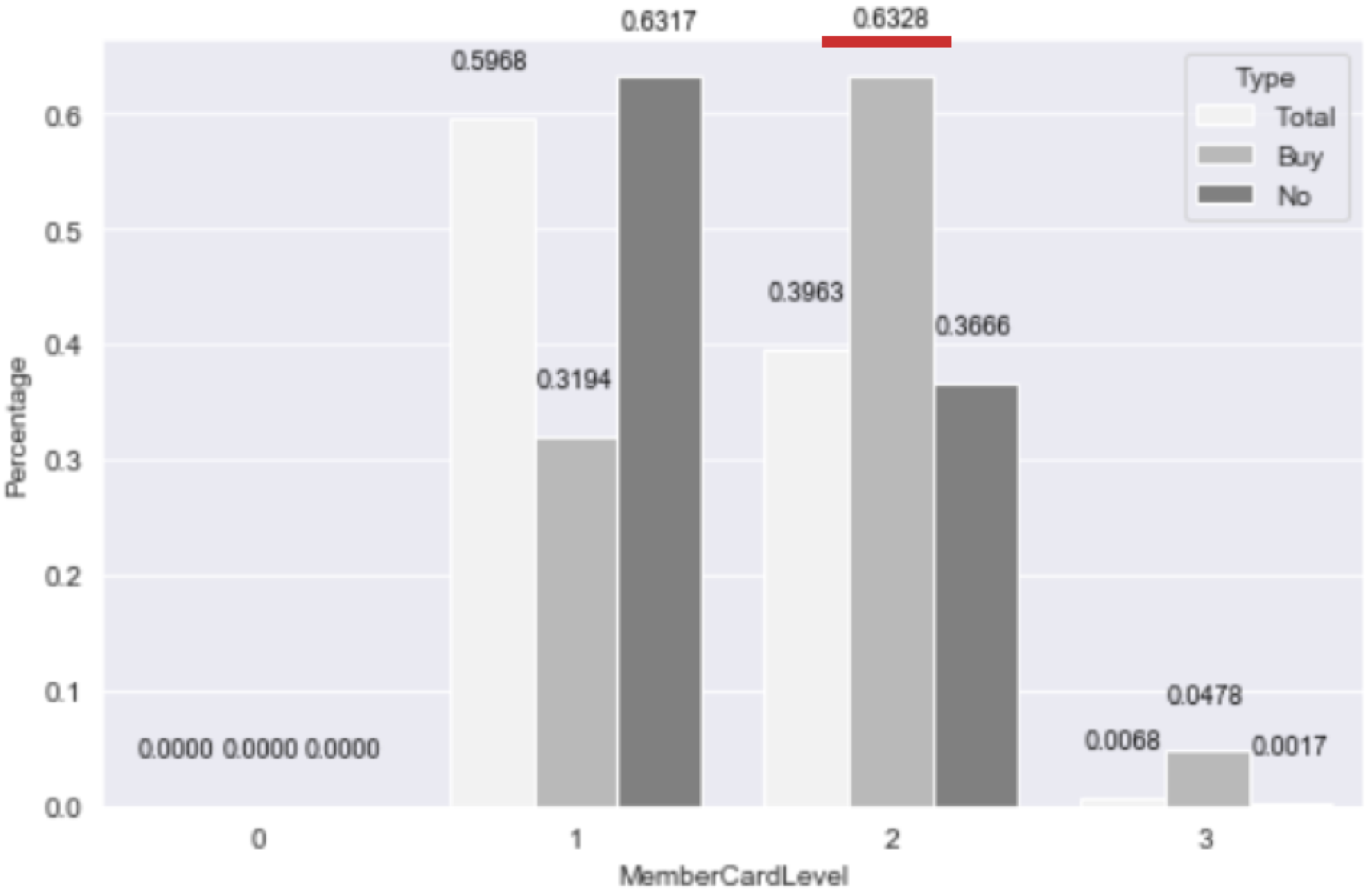
- 未來有購買行為的會員中，同意接收Email的族群的比例略低
- 未來無購買行為的會員中，同意接收Email的族群的比例略高



資料分析

MEMBER類別變數

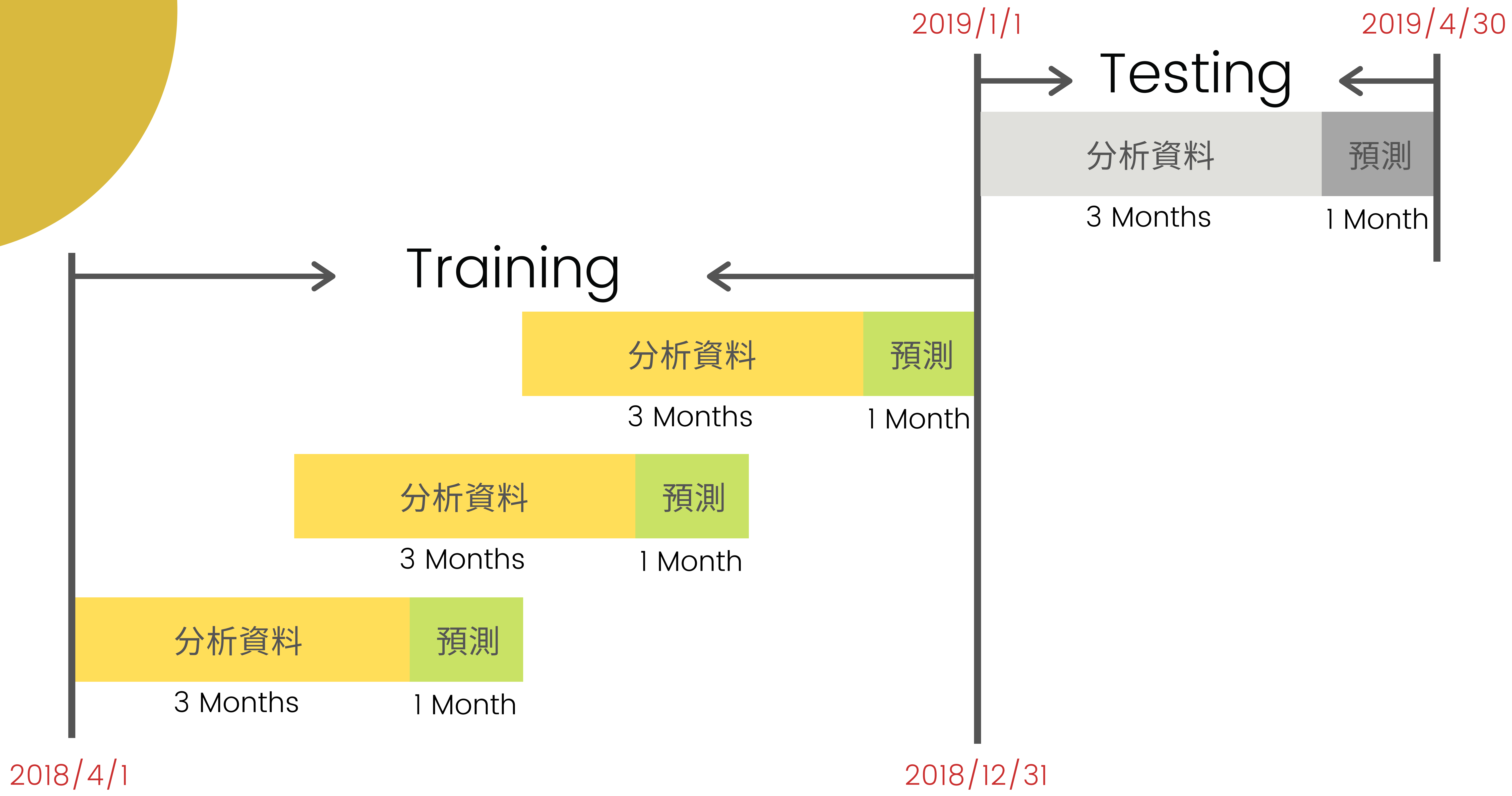
未來有購買行為的會員中，會員卡等級2的族群比例明顯較高



特徵選擇

69 FEATURES

Member	Order	Behavior
1 GENDER	11 PRESALESAMT	40 HITTIMES
2 AGE	12 PREBUYTIMES	41 SESSIONNUMBERCNT
3 ISONLINEMEMBER	13 SHOPPERIOD_MEAN	42 BEHAVIORTYPECNT_DIRECT
4 REGISTERSOURCETYPEDEF_IOSAPP	14 SHOPPERIOD_MAX	43 BEHAVIORTYPECNT_EMAIL
5 REGISTERSOURCETYPEDEF_ANDROIDAPP	15 SHOPPERIOD_MIN	44 BEHAVIORTYPECNT_FACEBOOK
6 REGISTERSOURCETYPEDEF_STORE	16 SHOPPERIOD_MEDIAM	45 BEHAVIORTYPECNT_GOOGLECPC
7 REGISTERSOURCETYPEDEF_LOCATIONWIZARD	17 PRESTATUSCNT_CANCEL	46 BEHAVIORTYPECNT_GOOGLEORGANIC
8 REGISTERSOURCETYPEDEF_WEB	18 PRESTATUSCNT_FAIL	47 BEHAVIORTYPECNT_INSTAGRAM
9 ISAPPINSTALLED	19 PRESTATUSCNT_FINISH	48 BEHAVIORTYPECNT_LINE
10 MEMBERCARDLEVEL	20 PRESTATUSCNT_NEW	49 BEHAVIORTYPECNT_LINESHOPPING
	21 PRESTATUSCNT_OVERDUE	50 BEHAVIORTYPECNT_OTHERS
	22 PRESTATUSCNT_RETURN	51 TRAFFICSOURCECATEGORYCNT_CART
	23 PRESTATUSCNT_SHIPPING	52 TRAFFICSOURCECATEGORYCNT_FAV
	24 PRECHANNELTYPECNT_MALL	53 TRAFFICSOURCECATEGORYCNT_PURCHASE
	25 PRECHANNELTYPECNT_OFFICIALECOM	54 TRAFFICSOURCECATEGORYCNT_SEARCH
	26 PRECHANNELTYPECNT_POS	55 TRAFFICSOURCECATEGORYCNT_VIEWSALEPAGE
	27 PREPAYMENTTYPECNT_ATM	56 TRAFFICSOURCECATEGORYCNT_VIEWSALEPAGECATE
	28 PREPAYMENTTYPECNT_CASH	57 SOURCETYPECNT_APP
	29 PREPAYMENTTYPECNT_CREDITCARDONCE	58 SOURCETYPECNT_WEB
	30 PREPAYMENTTYPECNT_FAMILY	59 OPERATIONSYSTEMCNT_ANDROID
	31 PREPAYMENTTYPECNT_LINEPAY	60 OPERATIONSYSTEMCNT_CHROMEOS
	32 PREPAYMENTTYPECNT_SEVENELEVEN	61 OPERATIONSYSTEMCNT_INTELMACOSX
	33 PRESHIPPINGTYPECNT_FAMILY	62 OPERATIONSYSTEMCNT_LINUX
	34 PRESHIPPINGTYPECNT_FAMILYPICKUP	63 OPERATIONSYSTEMCNT_UBUNTU
	35 PRESHIPPINGTYPECNT_HOME	64 OPERATIONSYSTEMCNT_WINDOWS
	36 PRESHIPPINGTYPECNT_LOCATIONPICKUP	65 OPERATIONSYSTEMCNT_IOS
	37 PRESHIPPINGTYPECNT_SEVENELEVEN	66 ONLINETIME_MEAN
	38 PRESHIPPINGTYPECNT_SEVENELEVENPICKUP	67 ONLINETIME_MAX
	39 PRESHIPPINGTYPECNT_STORE	68 ONLINETIME_MIN
		69 ONLINETIME_MEDIAM


2019/1/1
2019/4/30
Testing
分析資料
預測
3 Months
1 Month
Training
分析資料
預測
3 Months
1 Month
分析資料
預測
3 Months
1 Month
分析資料
預測
3 Months
1 Month
2018/4/1
2018/12/31

訓練/測試場景

訓練參數

訓練模型

Random Forest

K-FOLD

3

特徵數

69

訓練筆數

23852

Buy/NoBuy~= 4:1(過濾掉無法分析之後)

```
StratifiedKFold(n_splits=3, random_state=0, shuffle=False)
              precision    recall  f1-score   support

       NoBuy       0.45      0.17      0.25      1590
         Buy       0.82      0.95      0.88      6361

   micro avg       0.79      0.79      0.79      7951
   macro avg       0.64      0.56      0.56      7951
weighted avg       0.75      0.79      0.75      7951

Accuracy= 0.7924789334674883
              precision    recall  f1-score   support

       NoBuy       0.42      0.16      0.23      1590
         Buy       0.82      0.94      0.88      6361

   micro avg       0.79      0.79      0.79      7951
   macro avg       0.62      0.55      0.56      7951
weighted avg       0.74      0.79      0.75      7951

Accuracy= 0.7880769714501321
              precision    recall  f1-score   support

       NoBuy       0.55      0.27      0.36      1590
         Buy       0.84      0.94      0.89      6360

   micro avg       0.81      0.81      0.81      7950
   macro avg       0.70      0.61      0.63      7950
weighted avg       0.78      0.81      0.78      7950

Accuracy= 0.810188679245283
```

預測結果

	precision	recall	f1-score	support
NoBuy	0.98	0.56	0.72	33820
Buy	0.12	0.88	0.21	2335
micro avg	0.58	0.58	0.58	36155
macro avg	0.55	0.72	0.46	36155
weighted avg	0.93	0.58	0.68	36155

Tested accuracy = 0.5828239524270502

- **2019-4會員數 ~=36155**

  Buy/NoBuy~= 1:9, 過濾掉無法分析之後

- **預測前(假設全部投放)**

  36155

- **預測後**

  36155*42% =15185

- **提升58%經濟效益**

省下一半廣告費用

Follow_Ups

輔助NAPL一定可以提升預測的準確性

利用會員特徵的變化量產生新的特徵，提升準確率

針對購買會員做進一步分析

購買會員的變化程度
增加關鍵特徵的權重

Thank you!

THE END