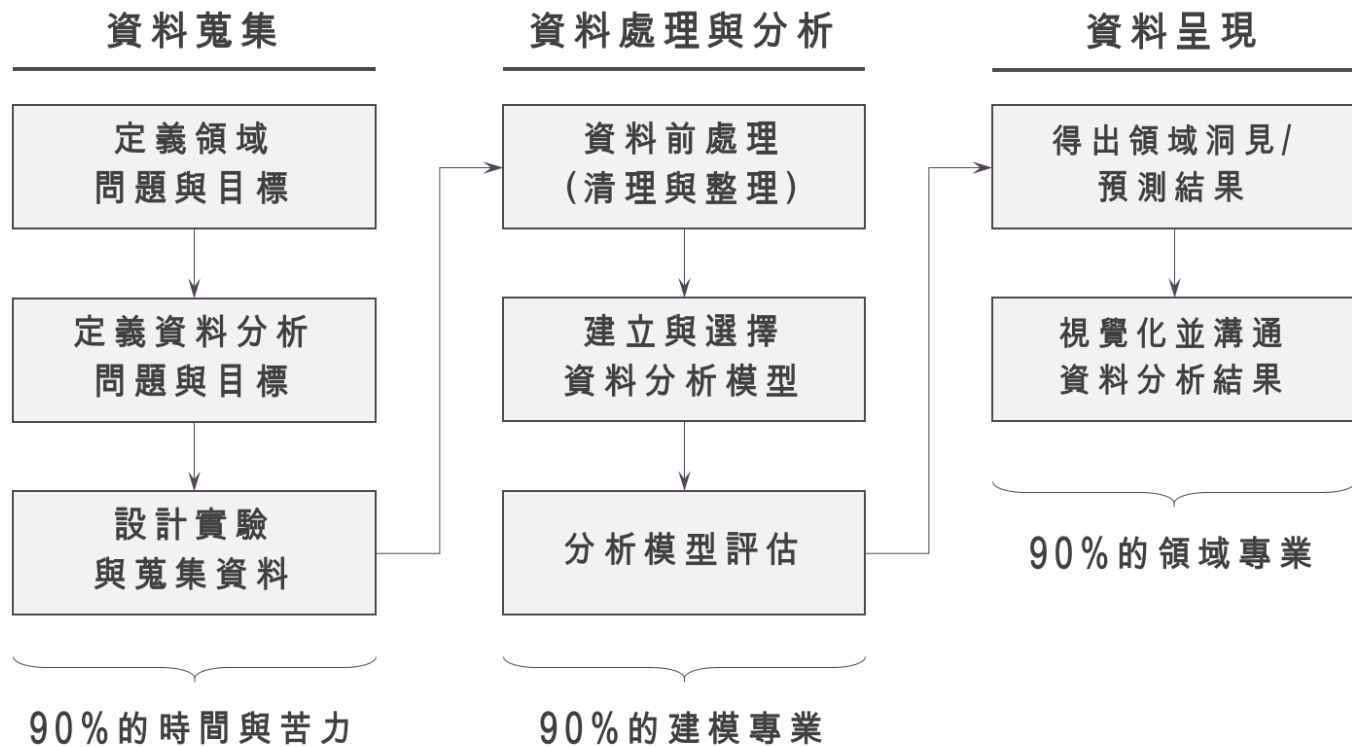


# 91APP 資料科學實務分享

莊竣丞 Zack Jhuang @ 91APP

# 資料專案的完整流程

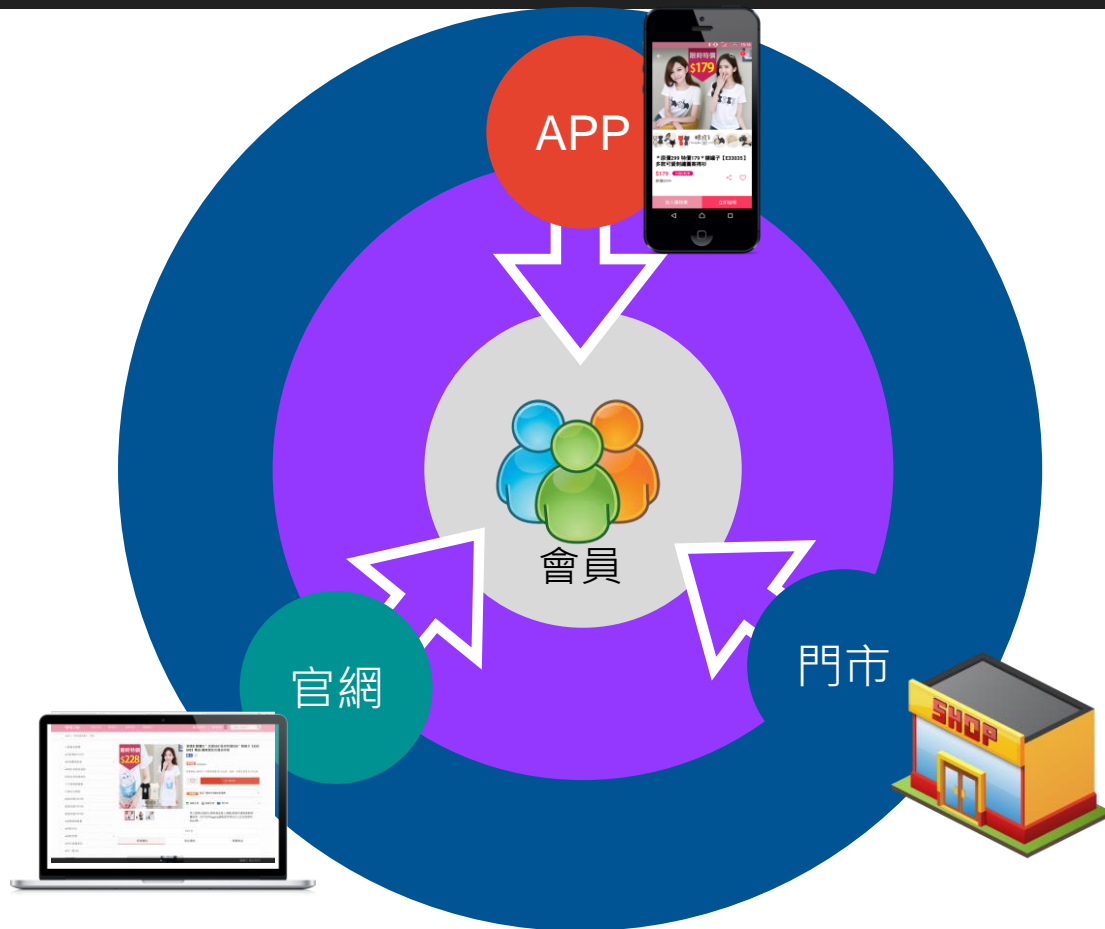


## 資料分析模型建立與評估：根據目標選擇方法

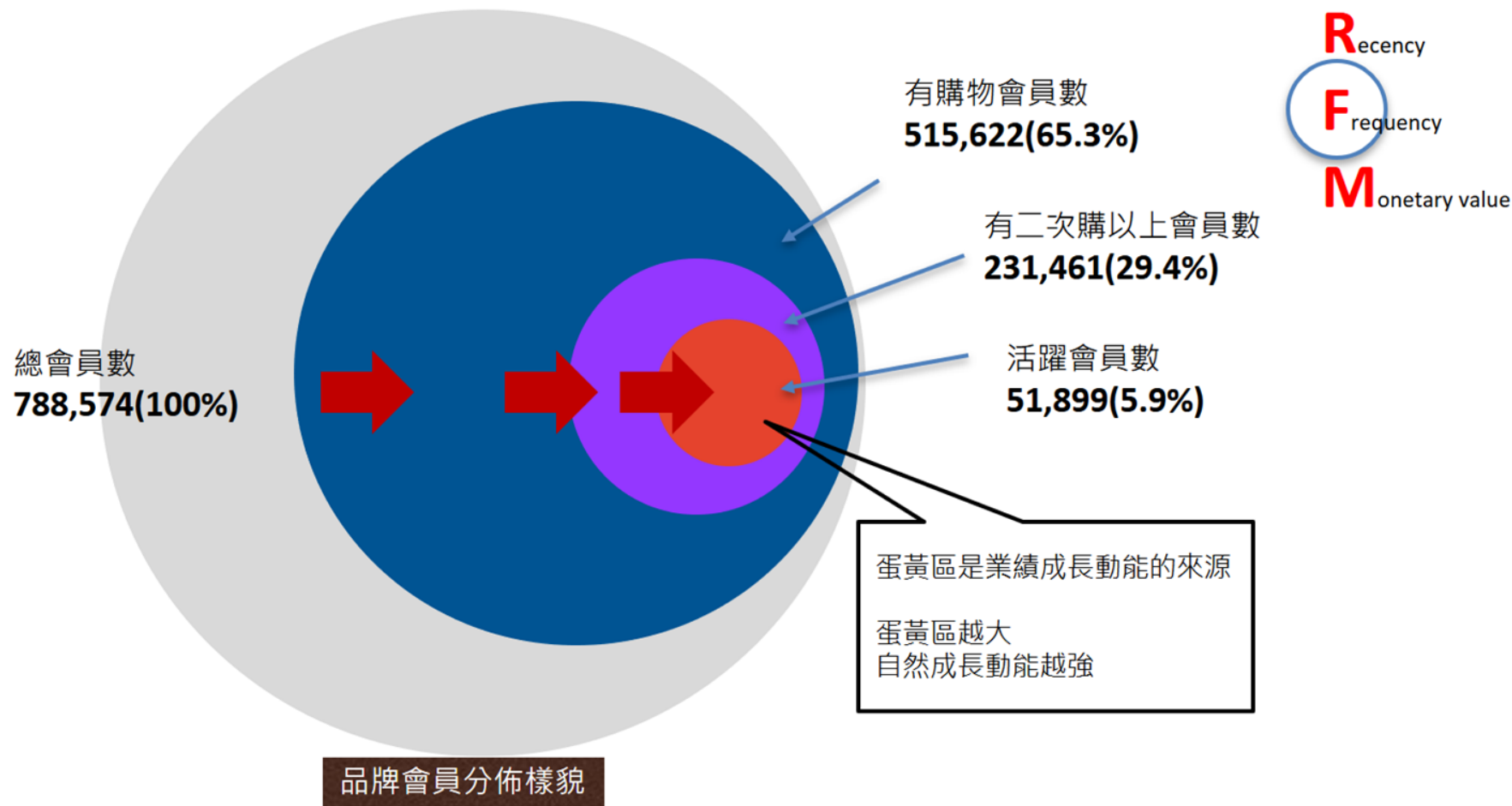
	商業建模	統計模型	機器學習
目標	<ul style="list-style-type: none"><li>了解市場現況作為決策分析的工具</li><li>在有限資料下進行市場預測</li></ul>	<ul style="list-style-type: none"><li>找出因果關係並驗證假說</li><li>進行結構性資料的個體或趨勢預測</li></ul>	<ul style="list-style-type: none"><li>進行準確的個體預測或個體分群</li><li>從非結構化的資料中找出重要的變數</li></ul>
強項	<ul style="list-style-type: none"><li>透過商業直覺建立符合需求的模型</li><li>容易執行且再利用</li></ul>	<ul style="list-style-type: none"><li>在雜訊中分離重要的訊號</li><li>經過科學驗證</li></ul>	<ul style="list-style-type: none"><li>自動且準確地進行個體預測或分群</li><li>規模化且彈性化</li></ul>
弱項	<ul style="list-style-type: none"><li>通常不會很準確</li><li>通常難以再利用</li></ul>	<ul style="list-style-type: none"><li>資料可信度非常重要</li><li>較深的理論知識</li></ul>	<ul style="list-style-type: none"><li>難以解釋解果</li><li>較深的理論知識</li></ul>

# 會員經營





# 會員經營的目標：提高新客與活躍會員數



低命中率

高行銷成本

100,000  
Members

Coupon

%

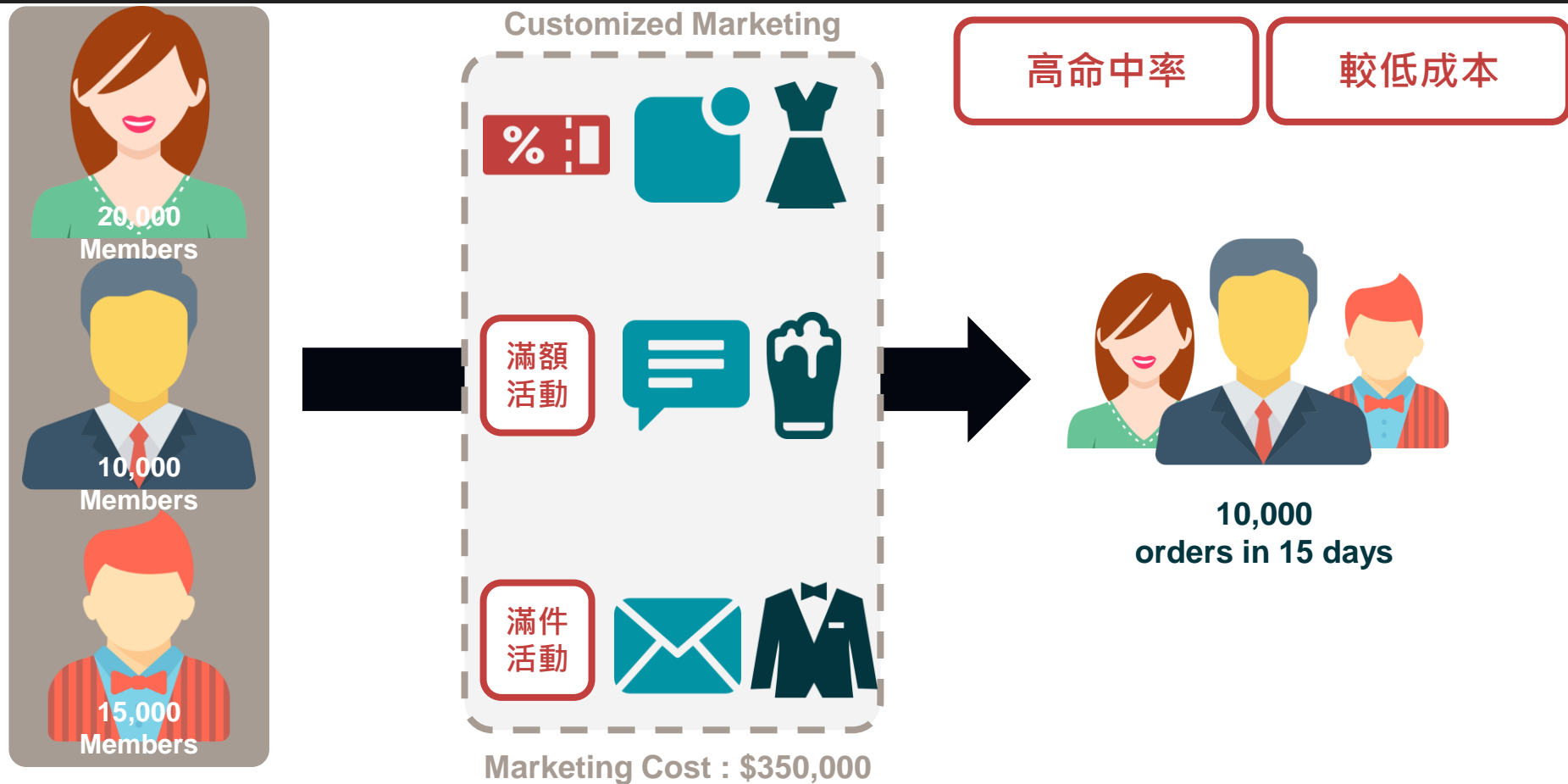
滿額  
活動

滿件  
優惠

EDM

Marketing Cost :  
\$1,000,000

5,000  
orders in 15 days





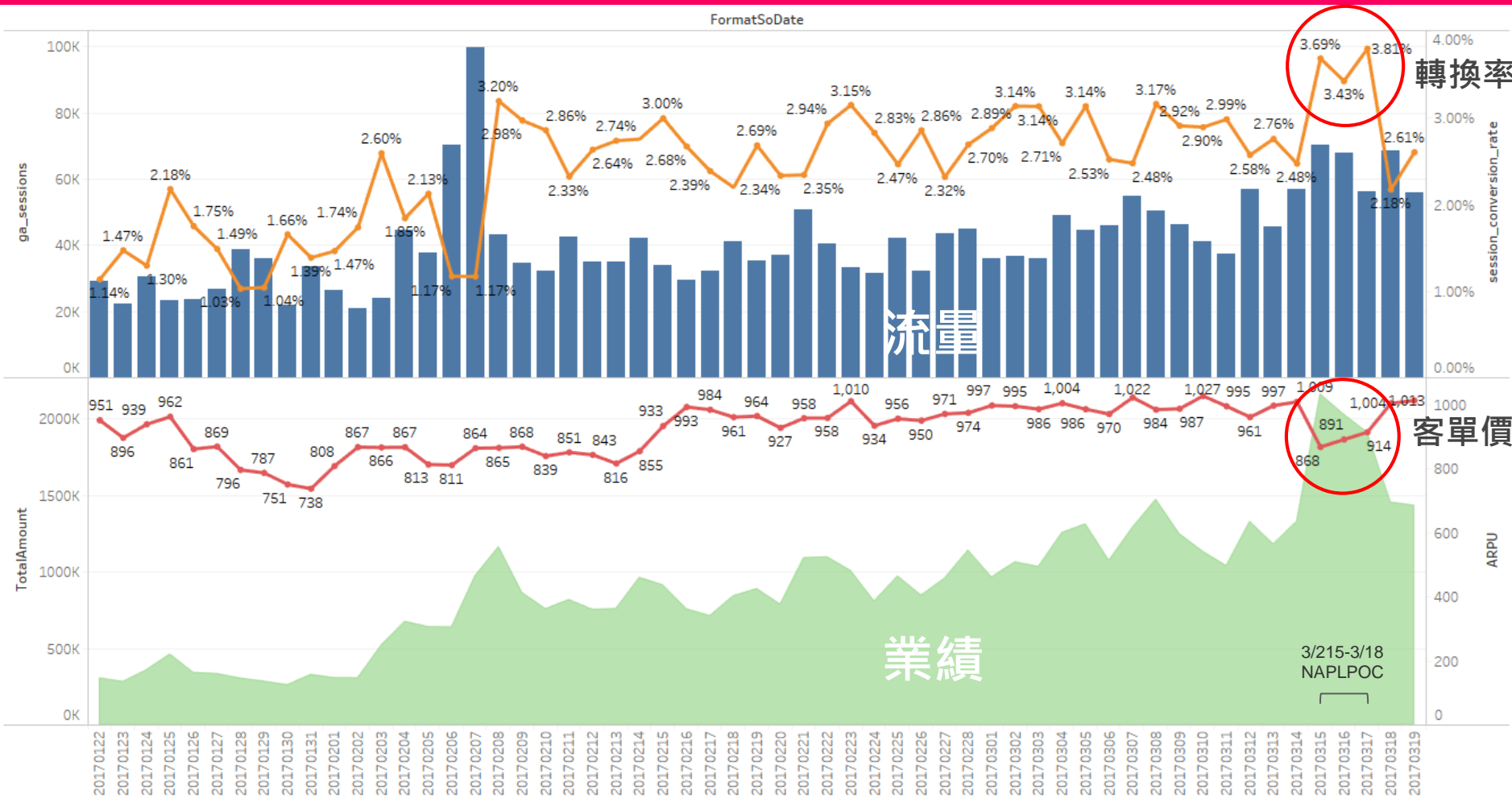
# NAPL 模型 - 將“有購物過”的會員分成 4 群



到達頁不同

實驗分組	溝通媒介	到達頁	未購	新客	活躍客	潛在客戶
實驗組A	推播	我的折價券	直接送你 100 元嚐試（已在您帳戶）看你遲疑好久，我想是價格吧？這活動不要錯過了 .a2 (折價券26746)	再送你 60元折價券嚐試第二次（結帳即可直接使用）雖然你才剛成為小麥客，但這次年前最後一次活動不要錯過！ .a4 (折價券26747)	老朋友就不多說了，再補貼你 25（給你用在年前最後一檔強品活動（\$\$老闆已偷塞到你帳戶裡啦） .a8 (折價券26750)	談～送你 4 元直接買（噓！已偷塞到你帳戶）不見了，年前這檔強品特價活動不要錯過 .a6 (折價券26748)
實驗組B	推播	活動頁公版	折價券金額高到低			
實驗組C	推播	小分類頁	直接送你 100 元嚐試（已在您帳戶）看你遲疑好久，我想是價格吧？這活動不要錯過了 .c2 (折價券26746)	（結帳即可直接使用）雖然你才剛成為小麥客，但這次年前最後一次活動不要錯過！ .c4 (折價券26747)	給你用在年前最後一檔強品活動（\$\$老闆已偷塞到你帳戶裡啦） .c8 (折價券26750)	談～送你 4 元直接買（噓！已偷塞到你帳戶）不見了，年前這檔強品特價活動不要錯過 .c6 (折價券26748)
實驗組D	推播	小分類頁	一致折價券金額			
實驗組E	推播	小分類頁	直接送你 25 元嚐試（已在您帳戶）看你遲疑好久，我想是價格吧？這活動不要錯過了 .e2 (折價券26754)	再送你 45元折價券嚐試第二次（結帳即可直接使用）雖然你才剛成為小麥客，但這次年前最後一次活動不要錯過！ .e4 (折價券26753)	老朋友就不多說了，再補貼你 30（給你用在年前最後一檔強品活動（\$\$老闆已偷塞到你帳戶裡啦） .e8 (折價券26751)	談～送你 4 元直接買（噓！已偷塞到你帳戶）不見了，年前這檔強品特價活動不要錯過 .e6 (折價券26748)
對照組	推播	小分類頁	純推播			

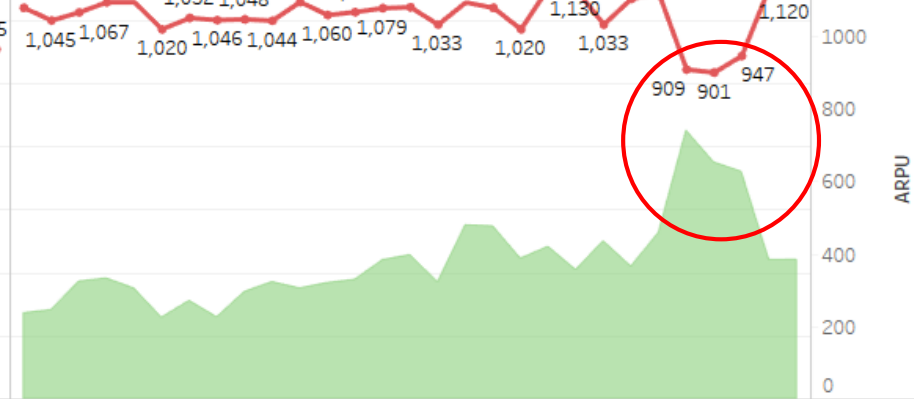
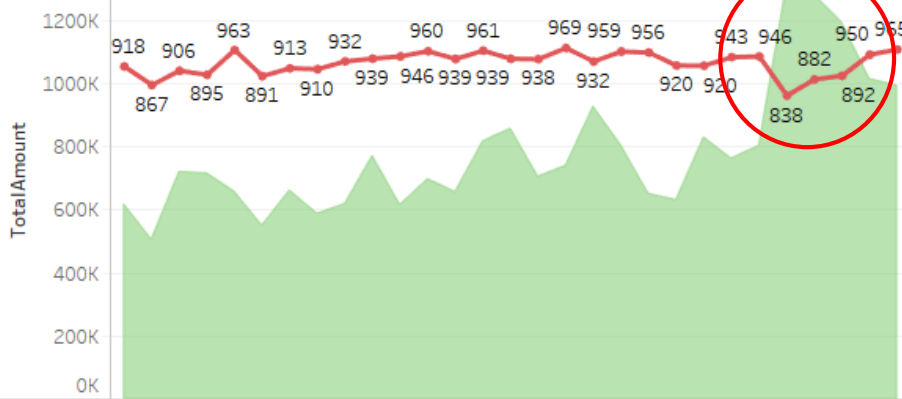
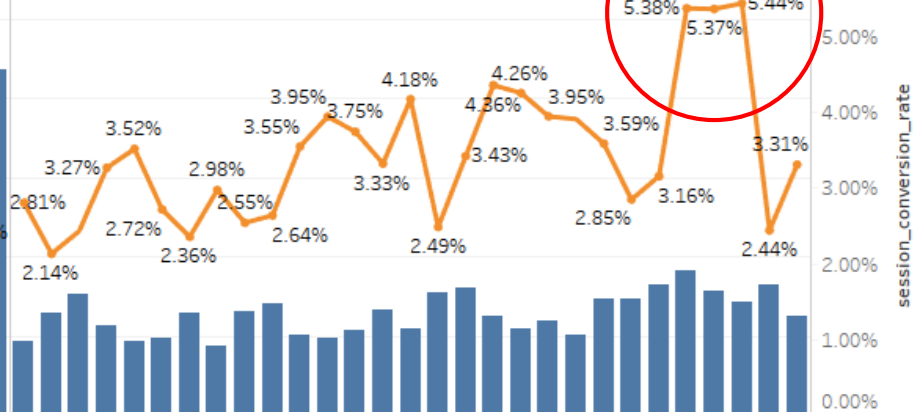
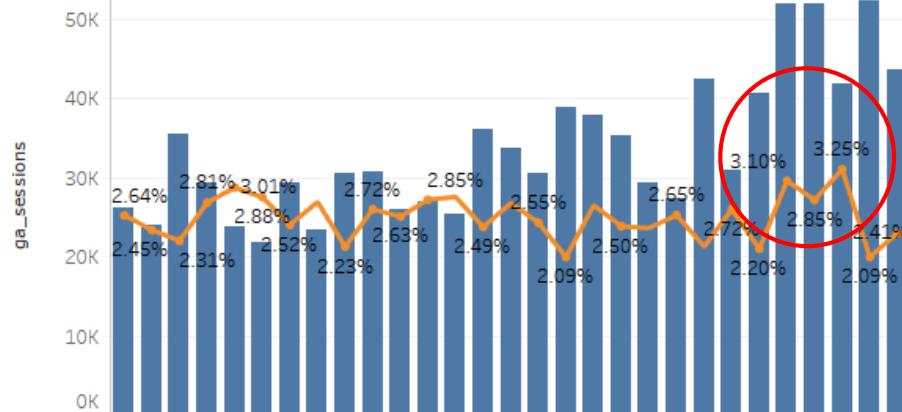
# 總體營運績效



官網 / APP 營運績效

官網

APP



比較各群回購率差異

行銷組合

實驗分組	未購				新客				活躍客				潛在客戶				總銷售額	轉換率
	名單數	訂單數	銷售額	轉換率	名單數	訂單數	銷售額	轉換率	名單數	訂單數	銷售額	轉換率	名單數	訂單數	銷售額	轉換率		
實驗組A	1,100	31	15,647	2.82%	4,000	125	80,718	3.13%	4,500	188	138,475	4.18%	1,400	38	26,822	2.71%	261,662	3.47%
實驗組B	1,100	27	17,675	2.45%	4,000	100	62,498	2.50%	4,500	179	132,296	3.98%	1,400	34	25,595	2.43%	238,064	3.09%
實驗組C	1,100	29	19,639	2.64%	4,000	105	68,475	2.63%	4,500	203	143,251	4.51%	1,400	39	27,421	2.79%	258,786	3.42%
實驗組D	1,100	13	9,044	1.18%	4,000	73	49,400	1.83%	4,500	211	146,537	4.69%	1,400	29	21,394	2.07%	226,375	2.96%
實驗組E	1,100	9	5,787	0.82%	4,000	109	72,182	2.73%	4,500	350	234,639	7.78%	1,400	34	25,483	2.43%	338,091	4.56%
對照組	1,100	7	4,228	0.64%	4,000	58	43,653	1.45%	4,500	167	134,617	3.71%	1,400	22	15,929	1.57%	198,427	2.31%
總計	6,600	116	72,020	1.76%	24,000	570	376,926	2.38%	27,000	1,298	929,815	4.81%	8,400	196	142,644	2.33%	1,521,405	3.30%

# SVM回購預測-模型特性

## Trade Off

抓得準，就抓不多  
抓得多，就抓不準

命中率

$$\text{Precision Rate} = \text{TP} / (\text{TP} + \text{FP})$$

抓取量

$$\text{Recall Rate} = \text{TP} / (\text{TP} + \text{FN})$$

		預測狀況	
		預測有回購	預測未回購
實際狀況	實際回購	True Positive(TP)	False Negative(FN)
	實際未回購	False Positive(FP)	True Negative(TN)

# SVM 實驗設計

實驗分組	預測名單	人數	溝通方式	折價券金額
A	Random	4,000	Facebook、 EDM、 簡訊、 推播	100元折價券
B	Active & Potential	4,000		
C	SVM(Frequency >= 2)	4,000		
D	SVM(Frequency >= 2)	4,000		無折價券
設計為期15天的活動，15天後觀察回購率檢視結果				

# SVM 成效衡量

預測名單	人數	折價券金額	購買人數	貢獻金額	轉換率
Random	4000	100元折價券	139	96,912	3.48%
Active & Potential	4000		318	243,887	7.95%
SVM(Frequency >= 2)	4000		646	521,937	16.15%
SVM(Frequency >= 2)	4000	無折價券	433	370,596	10.83%
設計為期15天的活動，15天後觀察回購率檢視結果					

- SVM (with coupon) 名單轉換率比 Active & Potential高兩倍
- SVM 名單受到行銷刺激後成長的幅度也比以往的經驗高



# 促銷機制



- 折扣活動 (Promotion)

- 任選優惠價
- 滿件打折
- 滿件折現
- 滿額打折
- 滿額折現
- 滿額贈

- 折價券活動 (Ecoupon)

- 首下載APP折價券 / 優惠券
- 輸入券號型
- 一般領取型
- 生日禮
- 開卡禮

- 滿額折現
- 滿額打折
- 滿件折現
- 滿件打折

### 新增折扣活動

\* 活動類型

☒ 現折 ☐ 任選優惠 ☐ 滿額贈

☐ 滿件折現

\* 活動名稱

(0/20)

滿件打折

滿額折現

滿額打折

滿件折現

滿件打折

- 一檔活動最多可設定 3 階折抵條件
  - 商品編號無法重複，一支商品一次僅能參加一檔活動。
- 例如，可同時設定：
  - 任選 2 件 399 元
  - 任選 3 件 599 元
  - 任選 4 件 799 元

均一價 ■ 699

🕒 2016/10/31 ~ 2017/01/31



均一價 ■ 799

🕒 2016/10/31 ~ 2017/01/31



均一價 ■ 999

🕒 2016/10/31 ~ 2017/01/31



- 消費者購買指定範圍的商品達門檻條件時，可以獲得指定贈品

## 滿1111贈💎我的心機黑珍珠潤白黑面膜

### • 活動辦法

2016/11/01 AM 10:00 ~ 2016/11/14 AM 11:00

歡慶1111購物嘉年華 滿💎1111送我的心機黑珍珠絲光潤白黑面膜乙片 跟著姊妹邊敷面膜邊逛Ann'S好開心~~~~~

### • 活動規則

消費滿額送贈品

詳細活動規則請參照活動辦法

商店保有修改活動內容之權利

數量有限，送完為止

### • 活動贈品



# 折價券活動 - 一般領取/輸入券號

## 一般領取

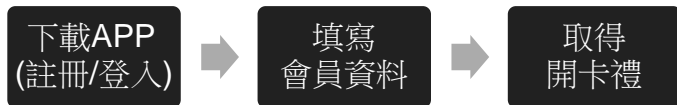
- 發送對象不限

## 輸入券號

- 單一序號，限量發送
  - 部落客操作
  - 粉絲頁操作
  - 分群會員操作
- 不同券號，限用一次
  - 會員轉移
  - 客服補發

# 開卡禮/生日禮 - (會員經營)

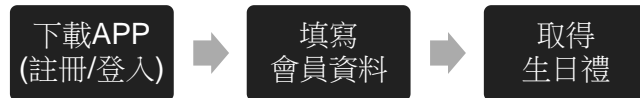
- 開卡禮



單一帳號只能領取一次

提供誘因讓會員寫資料，供未來分群使用

- 生日禮



每年只能領取一次

促使會員至少一年回來一次



# 折價券活動 - 首下載APP

## ● 導引消費者下載APP並消費





# 促銷機制的使用時機？

我想要消費者多買一點



滿件折扣、滿額折扣、滿額贈、  
任選優惠價、購物車加購

我想要刺激消費者購物但要能  
自由選購商品



一般領取型折價券  
輸入券號型折價券

我想針對會員經營操作



生日禮、開卡禮

我想讓我的消費者都有用我們  
的APP



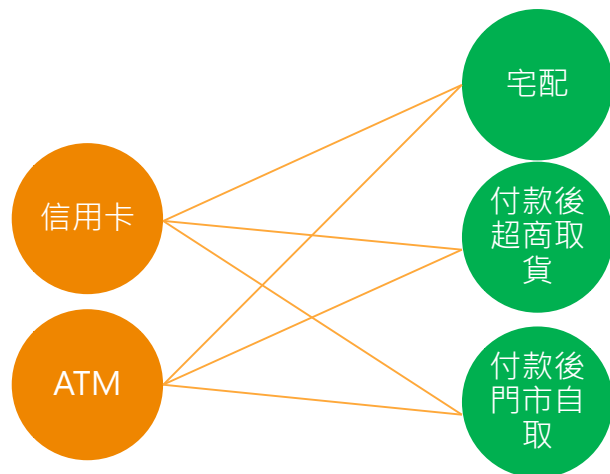
首下載折價券



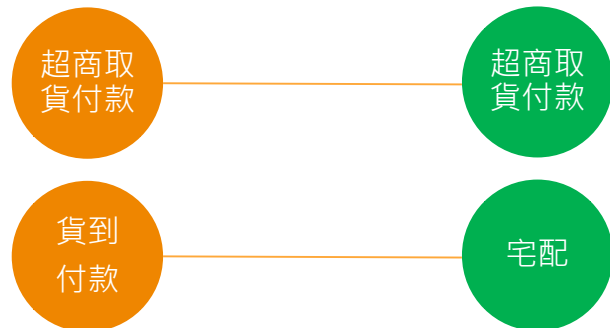
# 出貨管理 (金物流)



線上先付款  
再取貨



取貨付款  
同時



# DataSet 説明

# Order 資料表 - 1

欄位名稱	欄位定義
SalesOrderSlaveId	銷售明細序號 (可join CancelOrders)
DateId	日期
MemberId	會員編號(可join其他表)
ShopId	商家編號
SalePageId	商品頁編號
TradesOrderGroupCode	購物車編號 (同一次購買有相同購物車代碼，同一購物車下有多個 SalesOrderSlaveId)
TrackSourceTypeDef	流量來源：AndroidApp、IOSApp、Web
TrackDeviceTypeDef	使用裝置：Mobile、PC、Pad
LevelOneCategoryName	商品第一層級分類 (最概括分類)
LevelTwoCategoryName	商品第二層級分類

## Order 資料表 - 2

	欄位定義
LevelThreeCategoryName	商品第三層級分類
LevelFourCategoryName	商品第四層級分類
IsMajor	非贈品即為主件
IsGift	是否為買就送贈品 (1:是 ; 0:否)
Quantity	購買數量
UnitPrice	單品價格
PromotionDiscount	促銷活動折扣金額
EcouponDiscount	折價券折扣金額
SalesOrderSlaveTotalPayment	交易項目總金額 ( hint: 不是購物車或該次訂單總金額)
SalesOrderSlaveDateTime	訂單成立時間
SalesOrderSlaveStatusDef	訂單狀態： Cancel：取消訂單 Fail：訂單失敗 WaitingToShipping：等待運送中 Finish：訂單完成(消費者已拿到貨)

## Order 資料表 - 3

	欄位定義
PayProfileTypeDef	付款方式： ATM：Atm轉帳付款後出貨 SevenEleven：貨到付款7-11取貨 Family：貨到付款全家取貨 CreditCardOnce：信用卡付款後出貨
ShippingProfileTypeDef	取貨方式： SevenEleven：7-11取貨 Family：全家取貨 Home：宅配到家
EcouponId	折價券代碼(可對應到折價券Table)
CancelCauseDef	取消原因
CancelRefund	取消退款金額
CancelQuantity	取消數量
CancelOrderSlaveDateTime	取消時間
ReturnCauseDef	退貨原因
ReturnQuantity	退貨數量
ReturnRefund	退貨退款金額
ReturnGoodsOrderSlaveDateTime	退貨時間

# Member 資料表

欄位名稱	欄位定義	欄位名稱	欄位定義
ShopId	商店編號	DependentsTypeDef	是否有眷屬(小孩)
MemberId	會員編號	AnnualIncomeTypeDef	年收入
Recency	上一次購物間隔	RegisterDateTime	註冊日期
Frequency	購物周期內購物次數	RegisterSourceTypeDef	註冊管道： Web、AndroidApp、 iOSApp
Monetary	購買金額	MinOrderDate	第一次購買日期
Birthday	生日	IsInBlackList	是否在黑名單
GenderTypeDef	性別	DependentsTypeDef	是否有眷屬(小孩)
LocationState	地區	ProfessionTypeDef	職業
LocationCountry	國家	MaxAppOpenDateTime	最後開啟App日期
EducationTypeDef	教育程度： MasterDegreeOrAbove、 PrimaryOrBelow	MinAppOpenDateTime	第一次開啟App日期
ProfessionTypeDef	職業	MaxOrderDate	最後一次購買日期
MarialStatusTypeDef	婚姻狀態	LastBirthdayPresentYear	上一次生日禮領取年份



# Promotion 資料表

欄位名稱	欄位型態
PromotionId	活動序號
PromotionStartDateTime	活動開始日期
PromotionEndDateTime	活動結束日期
PromotionValidFlag	有效標誌
DiscountType	折扣活動類型： 任選優惠價 滿件打折 滿件折現 滿額打折 滿額折現 滿額贈

# Ecoupon 資料表

欄位名稱	欄位型態
EcouponId	折價券代碼
EcouponTypeDef	折價券類型： Drawout：一般領取型 Code：輸入券號 Birthday：生日禮 OpenCard：開卡禮 FirstDownload：
EcouponPurposeDef	目的
ECouponStartDateTime	開始使用日期
ECouponEndDateTime	結束使用日期
ECouponDiscountPrice	折價券面額
ECouponTotalQty	發行總數
ECouponQtyLimit	每張領取上限
ECouponTotalGiftLimit	贈送人數上限
ECouponsAppDrawOut	是否為 App 領取
ECouponsWebDrawOut	是否為 Web 站台領取
ECouponMaxDiscountLimit	最大折抵上限

# PromotionOrders 資料表

欄位名稱	欄位型態
PromotionId	活動序號
SalesOrderSlaveId	銷售明細序號
SalesOrderSlavePromotionDiscount	折扣金額
SalesOrderSlavePromotionDiscountRate	打折折數
SalesOrderSlavePromotionDiscountAllocated	折扣攤提金額
SalesOrderSlavePromotionDiscountAllocatedQty	折扣攤提商品件數
SalesOrderSlavePromotionValidFlag	生效標誌

# Tools for Homework

- Database
  - [Sqlserver](#)
  - MySQL / PostgreSQL (Google Cloud SQL)
- Data Manipulation
  - R / Python
  - [SQL](#)
- Visualization
  - Data Studio
  - Metabase
  - [Tableau Desktop](#)
  - Qlik Sense Desktop



SQLSERVER2014.AdventureWorks2012 - HumanResources.Employee - Microsoft SQL Server Management Studio (Administrator)

File Edit View Project Debug Table Designer Tools Window Help

Object Explorer

Connect

SQLSERVER2014 (SQL Server 12.0.1736 - SQLSERVER)

Databases

- System Databases
- Database Snapshots
- AdventureWorks2012
  - Database Diagrams
  - Tables
    - System Tables
    - FileTables
    - dbo.AWBuildVersion
    - dbo.DatabaseLog
    - dbo.ErrorLog
    - HumanResources.Department
    - HumanResources.Employee
      - Columns
        - BusinessEntityID (PK, FK, int, not null)
        - NationalIDNumber (nvarchar(15), not null)
        - LoginID (nvarchar(256), not null)
        - OrganizationNode (hierarchyid, not null)
        - OrganizationLevel (Computed Column, not null)
        - JobTitle (nvarchar(50), not null)
        - BirthDate (date, not null)
        - MaritalStatus (nchar(1), not null)
        - Gender (nchar(1), not null)
        - HireDate (date, not null)
        - SalariedFlag (Flag(bit), not null)
        - VacationHours (smallint, not null)
        - SickLeaveHours (smallint, not null)
        - CurrentFlag (Flag(bit), not null)

SQLSERVER2014.AdventureWorks2012.HumanResources.Employee

Column Name	Data Type	Allow Nulls
BusinessEntityID	int	<input type="checkbox"/>
NationalIDNumber	nvarchar(15)	<input type="checkbox"/>
LoginID	nvarchar(256)	<input type="checkbox"/>
OrganizationNode	hierarchyid	<input checked="" type="checkbox"/>
OrganizationLevel	Computed Column	<input checked="" type="checkbox"/>
JobTitle	nvarchar(50)	<input type="checkbox"/>
BirthDate	date	<input type="checkbox"/>
MaritalStatus	nchar(1)	<input type="checkbox"/>
Gender	nchar(1)	<input type="checkbox"/>
HireDate	date	<input type="checkbox"/>
SalariedFlag	Flag:bit	<input type="checkbox"/>
VacationHours	smallint	<input type="checkbox"/>

Properties

[Tb] HumanResources.Employee

(Identity)

(Name) Employee

Database Name AdventureWorks2012

Description Employee information

Schema HumanResources

Server Name sqlserver2014

Table Designer

Identity Column

Indexable Yes

Lock Escalation Table

Regular Data Space PRIMARY

Replicated No

Row GUID Column rowguid

Text/Image Filegroup PRIMARY

Column Properties

(General)

(Name) BusinessEntityID

Allow Nulls No

Data Type int

Default Value or Binding

Table Designer

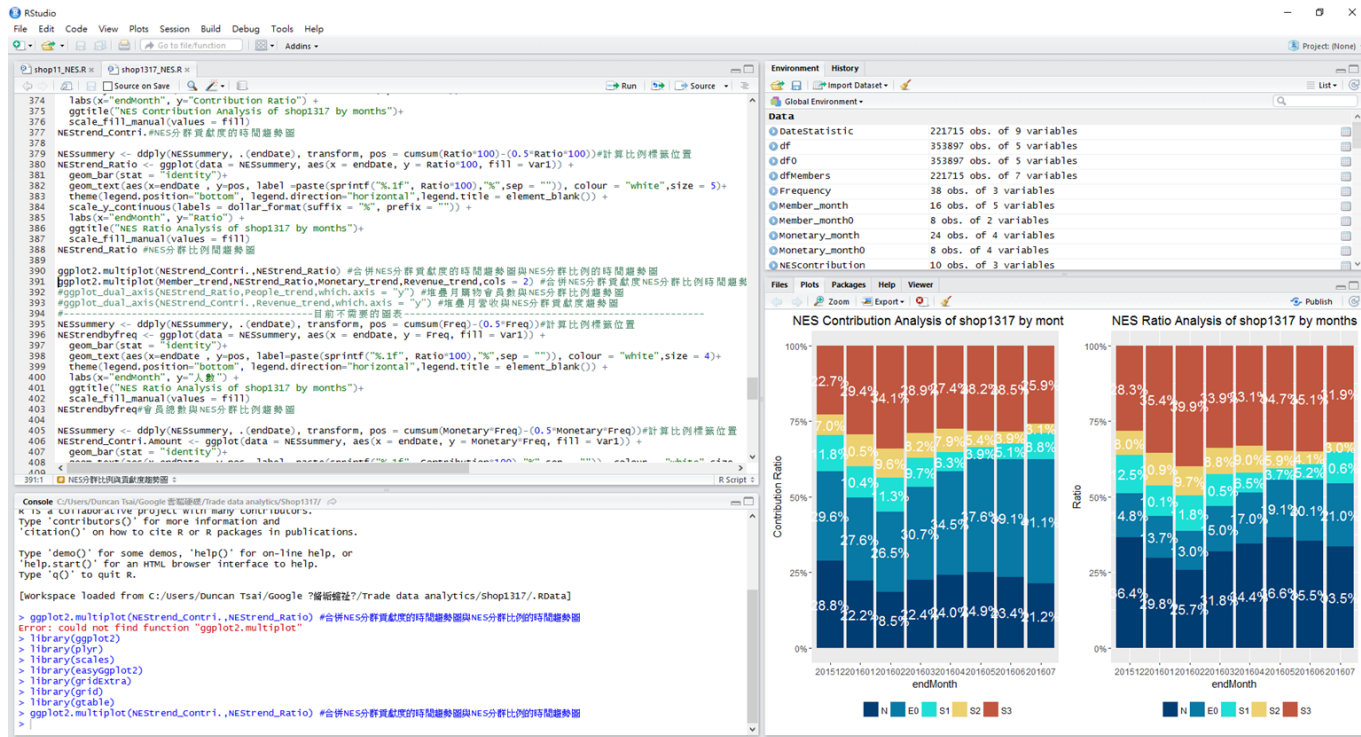
(General)

(Identity)



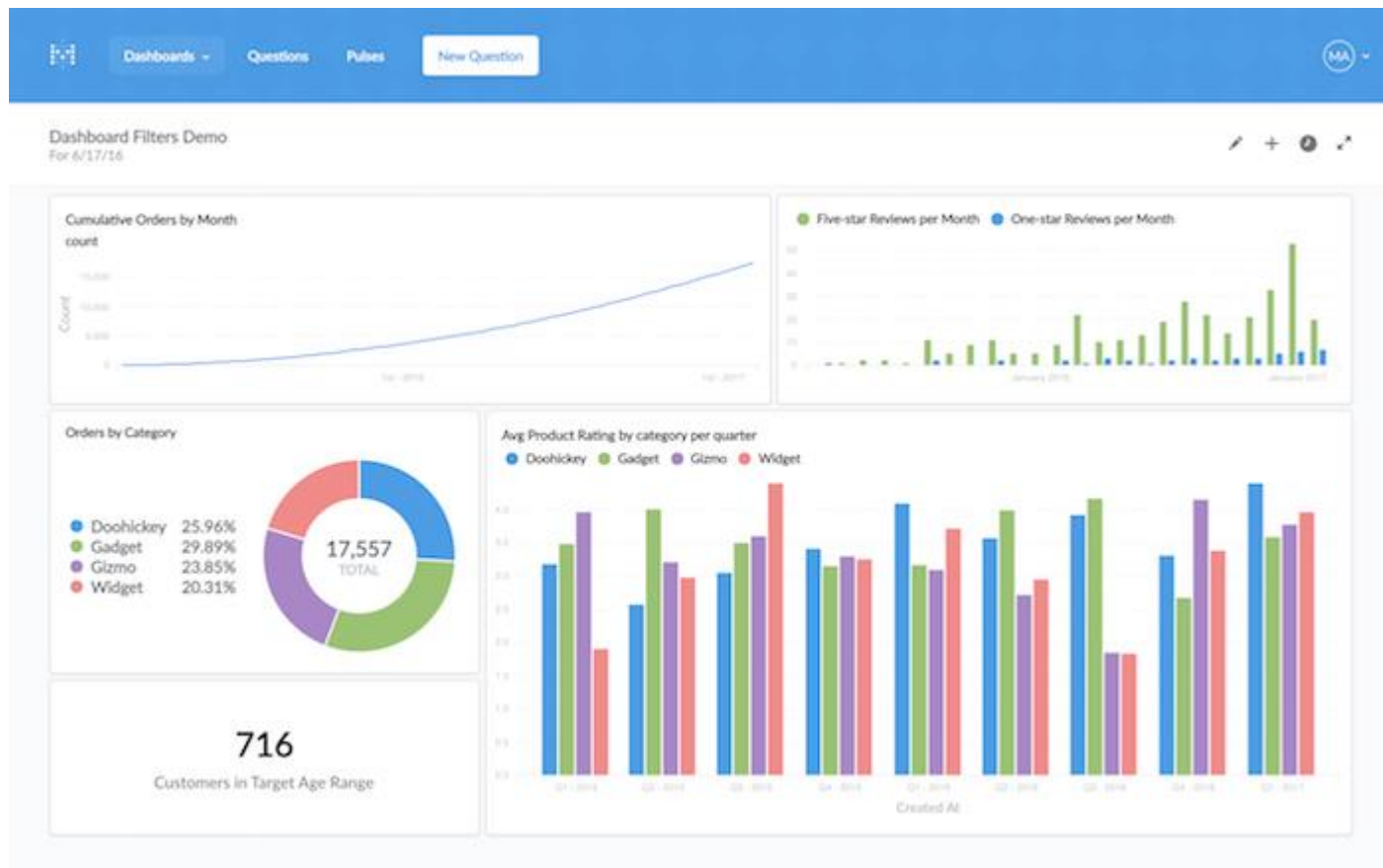
## Statistical computing and graphics

## Data Manipulation & Calculation





Metabase



### Overview

#### Click Through Rate & Cost

by Clicks, CTR, and Impressions

Clicks  
**10.7K**  
↑ 33.2%

CTR  
**2.33%**  
↓ -9.4%

Impressions  
**460K**  
↑ 47.0%



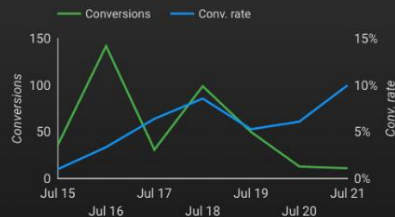
#### Conversion Rate & Cost

by Conversions Rate and Cost / Conv.

Conversions  
**382**  
↓ -35.3%

Conv. rate  
**3.57%**  
↓ -51.4%

Cost / Conv.  
**\$4.5**  
↑ 24.1%



#### Cost Per Click

by Cost, CPC, and CPM

Cost  
**\$1.7K**  
↓ -19.6%

Avg. CPC  
**\$0.2**  
↓ -39.7%

Avg. CPM  
**\$3.7**  
↓ -45.3%



#### Top Campaigns

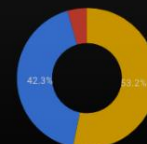
by CTR, Avg. CPC, and Cost / Conv.

	Campaign	Campaign ID	CTR	Avg. CPC	Cost / Conv.
1.	Campaign - T-Shirts	458390433	4.21%	\$0.14	\$6.26
2.	Campaign - Accessories	301191558	2.72%	\$0.08	\$0.82
3.	Campaign - Wearables	272290758	1.64%	\$0.34	\$6.34
4.	Campaign - Biking	235343532	1.52%	\$0.07	\$0.77
5.	Campaign - Dynamic Search Ads	272290518	0.84%	\$0.23	\$3.58
6.	Campaign - Active Wear	235343532	0.27%	\$0.37	\$6.08

#### Device Breakdown

by Clicks, Cost, and Conversions

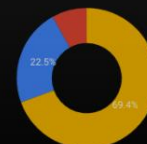
Mobile phones Tablets Computers



Clicks



Cost



Conversions



