

# 91APP X 大數據與商業分析課程

## DataSet 說明文件

91APP 數據服務組 CocoChung

2020/05/27

# 資料簡述

## 業態 - 時尚流行女鞋 (1間商家)

- **Member 會員資料**

大小: 329,871 rows × 10 columns

時間: ~2020/04/30

- **Order 交易資料 (主單資料)**  
**OrderSlave (子單資料)**

大小: 329,871 rows × 10 columns

時間: 2013/09/30 ~ 2020/04/30

- **Behavior 行為資料**

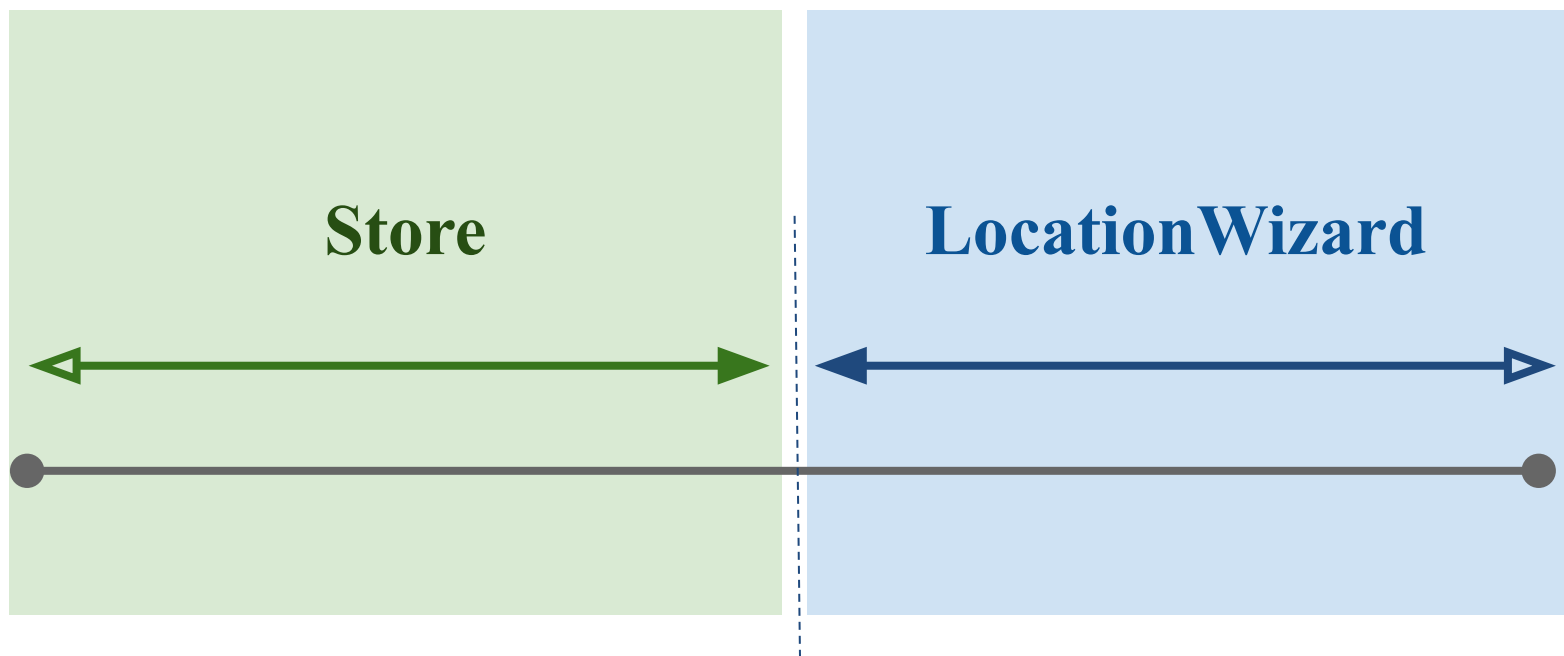
時間: 2018/06/06 ~ 2020/04/30



# 會員資料 MemberData

欄位名稱	欄位定義	格式	欄位解釋
MemberID	會員識別ID	Str	含英文 數字
RegisterSourceTypeDef	註冊來源	Str	<b>Store</b> ：門市註冊 <b>LocationWizard</b> ：小幫手註冊 <b>iOSApp</b> ：iOSApp註冊 <b>AndroidApp</b> ：AndroidApp註冊 <b>Web</b> ：官網註冊 <b>NaN</b> ：缺失值
RegisterDateTime	註冊日期	Datetime	YYYY / MM / DD hh:mm:ss
Gender	性別	Str	<b>Male</b> ：男 <b>Female</b> ：女 <b>NaN</b> ：缺失值
Birthday	生日	Date	YYYY / MM / DD 1900 / 01 / 01 為預設值
IsAppInstalled	是否下載APP	Str	True / False
IsEnableEmail	是否同意接收Email	Str	True / False
IsEnablePushNotification	是否同意接收Push	Str	True / False
IsEnableShortMessage	是否同意接收SMS	Str	True / False
MemberCardLevel	會員卡等級	Int	<b>最高等級</b> :40 <b>次高等級</b> :30 <b>中間等級</b> :20 <b>一般等級</b> :10 <b>缺失值</b> :0
Eland_uuid	意藍識別ID	Str	含英文 數字

# Store vs Location Wizard



導入門市小幫手

# 交易主單資料 OrderData

欄位名稱	欄位定義	格式	欄位解釋
MemberID	會員識別ID	Str	含英文 數字
TradesGroupCode	訂單編號	Str	主單編號 (含英文 數字)
TradesDateTime	交易日期	Datetime	YYYY / MM / DD hh:mm:ss
ChannelType	交易通路類型(大)	Str	OfficialECom：線上 Pos：線下 Mall：舊版商城 (2016年前) LocationWizard：線下小幫手打單
ChannelDetail	交易通路類型(小)	Str	DesktopOfficialWeb：官網 MobileWeb：手機版官網 iOSApp / AndroidApp：App 其他：線下櫃位/門市代碼
PaymentType	付款類型	Str	Cash：現金 ATM：ATM轉帳 Family：全家貨到付款 SevenEleven：7-11貨到付款 CreditCardOnce：信用卡一次付清 CreditCardInstallment：信用卡分期付款 LinePay：LinePay JKOPay：街口支付

# 將交易日期UTC時間轉為台灣時區

```
In [17]: 1 import pandas as pd
          2 from datetime import timedelta
```

```
In [20]: 1 raw_data = order[['DateId', 'TradesDateTime', 'TradesGroupCode']]
          2 raw_data.head(3)
```

Out[20]:

	DateId	TradesDateTime	TradesGroupCode
0	20110829	2011-08-29 10:06:25 UTC	C108008869
1	20110826	2011-08-26 12:17:21 UTC	C108007980
2	20110825	2011-08-25 13:53:06 UTC	C108007715

擷取所需欄位前19位  
(2011-08-29 10:06:25 含空格共19位)  
轉為時間格式後+8小時

```
In [11]: 1 raw_data['TradesDateTime'] = pd.to_datetime(raw_data['TradesDateTime'].astype(str).str[0:19]
          2 , format="%Y-%m-%d %H:%M:%S") + timedelta(hours = 8)
          3 raw_data.head(5)
```

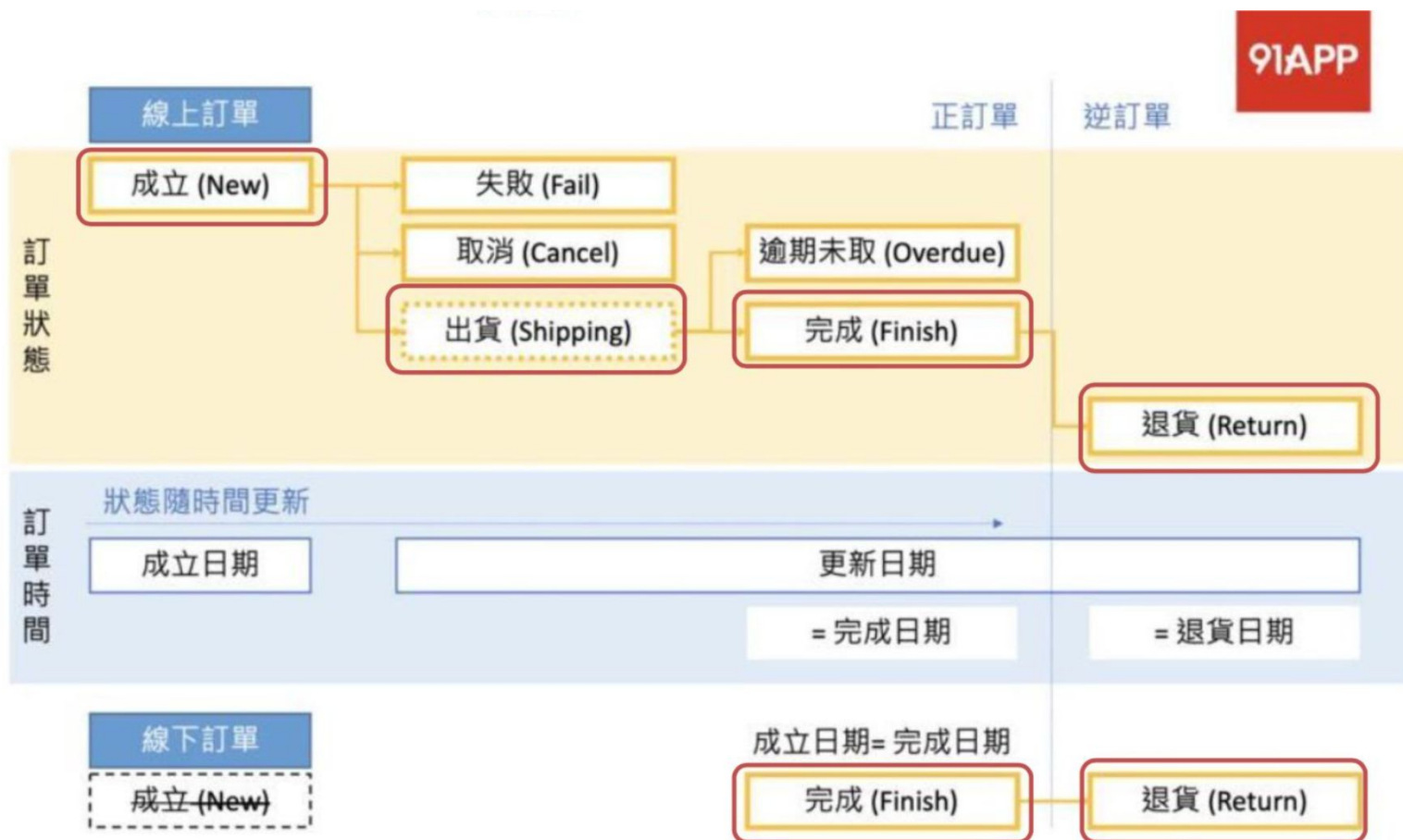
Out[22]:

	DateId	TradesDateTime	TradesGroupCode
0	20110829	2011-08-30 02:06:25	C108008869
1	20110826	2011-08-27 04:17:21	C108007980
2	20110825	2011-08-26 05:53:06	C108007715



欄位名稱	欄位定義	格式	欄位解釋
ShippingType	運送類型	Str	<b>Store</b> : 門市購買 <b>LocationPickup</b> : 線下櫃位/門市取貨 <b>Family</b> : 全家取貨付款 <b>FamilyPickup</b> : 全家純取貨 <b>SevenEleven</b> : 7-11取貨付款 <b>SevenElevenPickup</b> : 7-11純取貨 <b>Home</b> : 宅配至家裡
TsCount	訂單交易品類數	Int	訂單內總品類數量
Qty	訂單交易數量	Int	訂單交易數量(退貨Return: 負值)
TotalSalesAmount	實付金額	Int	已扣除折價和運費
TotalPrice	應付金額	Int	未扣除折價和運費
TotalDiscount	總折扣金額	Int	所有折扣總和
TotalPromotionDiscount	總折價優惠金額	Int	促銷活動折價
TotalCouponDiscount	總折價券優惠金額	Int	使用折價券折價
TotalLoyaltyDiscount	總會員折價優惠金額	Int	會員身分折價

欄位名稱	欄位定義	格式	欄位解釋
Status	訂單狀態	Str	<b>New</b> : 新訂單 <b>Fail</b> : 失敗訂單(未扣款) <b>Cancel</b> : 取消訂單 <b>Return</b> : 退貨 <b>Overdue</b> : 逾期未取 <b>Shipping</b> : 運貨中 <b>Finish</b> : 完成訂單



# 合併相同MemberID的訂單與會員資料

```
In [54]: 1 df1 = order[order['MemberID'] == 'U99462295'][['MemberID', 'TransactionNum', 'TradesDate']]
          2 df1
```

```
Out[54]:
```

	MemberID	TransactionNum	TradesDate
4	U99462295	zVBGf/pLjTKHQFQUEQ5zUTvS8Yp=	2018/7/8

```
In [60]: 1 df2 = member[member['MemberID'] == 'U2131156 '][['MemberID', 'Gender', 'Birthday']]
          2 df2
```

```
Out[60]:
```

	MemberID	Gender	Birthday
--	----------	--------	----------

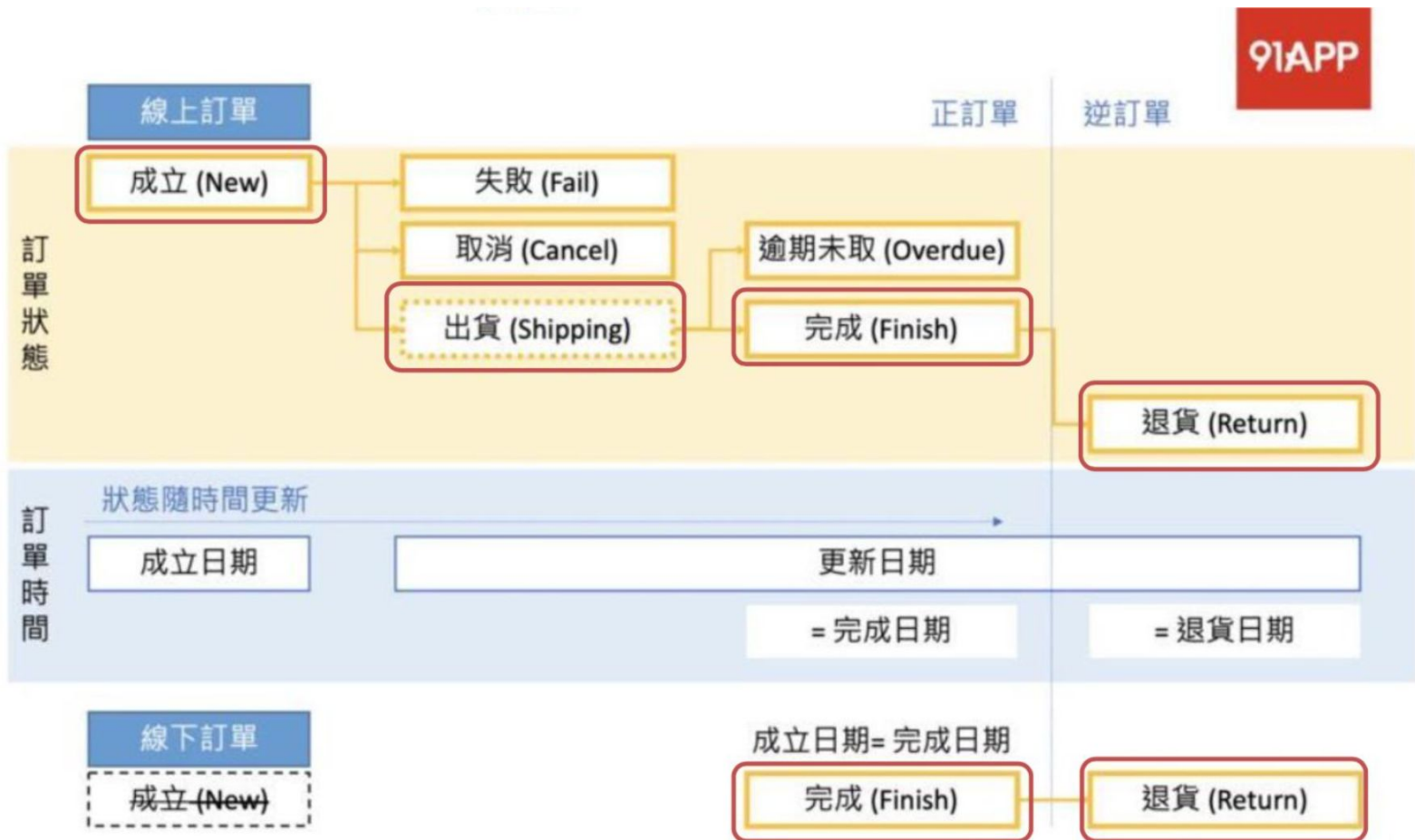
```
In [58]: 1 temp = pd.merge(df1, df2, on = 'MemberID', how = 'left')
          2 temp
```

# 交易子單資料 OrderSlaveData

欄位名稱	欄位定義	格式	欄位解釋
MemberID	會員識別ID	Str	含英文 數字
TradesGroupCode	訂單編號	Str	主單編號 (含英文 數字)
TradesDateTime	交易日期	Datetime	YYYY / MM / DD hh:mm:ss
ChannelType	交易通路類型(大)	Str	<b>OfficialECom</b> : 線上 <b>Pos</b> : 線下 <b>Mall</b> : 舊版商城 (2016年前) <b>LocationWizard</b> : 線下小幫手打單
ChannelDetail	交易通路類型(小)	Str	<b>DesktopOfficialWeb</b> : 官網 <b>MobileWeb</b> : 手機版官網 <b>iOSApp / AndroidApp</b> : App <b>其他</b> : 線下櫃位/門市代碼
PaymentType	付款類型	Str	<b>Cash</b> : 現金 <b>ATM</b> : ATM轉帳 <b>Family</b> : 全家貨到付款 <b>SevenEleven</b> : 7-11貨到付款 <b>CreditCardOnce</b> : 信用卡一次付清 <b>CreditCardInstallment</b> : 信用卡分期付款 <b>LinePay</b> : LinePay <b>JKOPay</b> : 街口支付 <b>空值</b> : 缺失值

欄位名稱	欄位定義	格式	欄位解釋
ShippingType	運送類型	Str	<b>Store</b> :門市購買 <b>LocationPickup</b> :線下櫃位/門市取貨 <b>Family</b> :全家取貨付款 <b>FamilyPickup</b> :全家純取貨 <b>SevenEleven</b> :7-11取貨付款 <b>SevenElevenPickup</b> :7-11純取貨 <b>Home</b> :宅配至家裡 <b>空值</b> :缺失值
OuterProductSkuCode	品牌商品編號	Str	訂單內總品類數量
ProductSkuCode	91app商品編號	Int	<b>空值</b> :缺失值
SalePageCode	線上商品頁面編號	Int	<b>空值</b> :缺失值
Qty	訂單交易數量	Int	訂單交易數量(退貨Return:負值)
TotalSalesAmount	實付金額	Int	已扣除折價和運費
TotalPrice	應付金額	Int	未扣除折價和運費
TotalDiscount	總折扣金額	Int	所有折扣總和

欄位名稱	欄位定義	格式	欄位解釋
Status	訂單狀態	Str	<b>New</b> : 新訂單 <b>Fail</b> : 失敗訂單 <b>Cancel</b> : 取消訂單 <b>Return</b> : 退貨 <b>Overdue</b> : 逾期未取 <b>Shipping</b> : 運貨中 <b>Finish</b> : 完成訂單



# 行為資料 BehaviorData



## 商店官網+APP 所蒐集之使用者行為相關資料

- User Registration (會員註冊)
- Main PageView (瀏覽首頁)
- Product PageView (瀏覽商品頁)
- Category PageView (瀏覽分類頁)
- Activity PageView (瀏覽活動頁)
- Search (搜尋)
- Add To Cart (加入購物車)
- Checkout (開始結帳)
- Purchase (購買)
- Traffic Source (流量來源)

欄位名稱	欄位定義	格式	欄位解釋
mid	91商家編號	Int	此次資料集商家編號為:123
dc	瀏覽裝置	Str	<b>Web</b> : 網頁瀏覽 <b>Android</b> :Android瀏覽 <b>iOS</b> :iOS瀏覽
did	廣告識別碼	Str	<b>AAID(Android)</b> <b>IDFA(iOS)</b> <b>Web DeviceId</b> XXXXXXXX-XXXX-XXXX-XXXX-XXXXXXXXXXXX
uid	會員識別ID	Str	<b>MemberID</b> <b>空值</b> :無對應會員
ht	時間戳記	Int	Unix Timestamp 格式(13碼)
ul	使用者語系	Str	<b>語言_地區(語言_script_地區)</b> <b>en</b> :英語 <b>en-HK</b> : 英語-香港 <b>en-SG</b> : 英語-新加坡 <b>en-tw</b> :英語-台灣 <b>en-us</b> : 英語-美國 <b>ja-jp</b> : 日語-日本 <b>zh-Hant</b> : 繁體中文 <b>zh-Hant-TW</b> : 繁體中文-台灣 <b>zh-TW</b> : 台灣

# Unix TimeStamp

## UnixTime

UNIX或類UNIX系統使用的時間表示方式

UTC1970年1月1日0時0分0秒起至現在的總秒數

```
1 import datetime
2 timestamp = datetime.datetime.fromtimestamp(1284101485)
3 print(timestamp.strftime('%Y-%m-%d %H:%M:%S'))
```

2010-09-10 14:51:25

本次資料集使用Unix Timestamp 格式(13碼)

```
1 import datetime
2
3 timestamp = "1523126888080"
4 your_dt = datetime.datetime.fromtimestamp(int(timestamp)/1000)
5 print(your_dt.strftime('%Y-%m-%d %H:%M:%S'))
```




2018-04-08 02:48:08

欄位名稱	欄位定義	格式	欄位解釋
geoid	地理識別ID	Str	TW: 台灣
bh	使用者行為	Str	<b>viewactivity</b> : 瀏覽活動頁面 <b>add</b> : 加入購物車 <b>viewcategory</b> : 瀏覽分類頁面 <b>checkout</b> : 開始結帳 <b>viewmainpage</b> : 瀏覽首頁 <b>viewproduct</b> : 瀏覽商品頁 <b>purchase</b> : 購買 <b>search</b> : 搜尋 <b>traffic</b> : 流量來源 <b>register</b> : 會員註冊
pr	使用者行為參數	Str	有中文 數字
pr1id	商品ID	Int	商品識別ID
pr1pr	商品單價	Float	商品各一單價
pr1qt	商品數量	Int	商品的數量
cu	貨幣代碼	Str	TWD: 台幣
tr	交易金額	Float	交易花費金額
ti	交易編號	Str	交易主單編號

欄位名稱	欄位定義	格式	欄位解釋
cs	廣告活動來源	Str	<b>Google</b> :Google 進入 <b>Line</b> :Line進入 <b>SMS</b> :簡訊進入 <b>FBAD</b> :FB廣告進入 <b>IGPO</b> :IGPO文進入
cm	廣告活動媒介	Str	<b>cpc</b> : Google Ads <b>organic</b> : 自然搜尋引擎 <b>referral</b> : 其他網站推薦連結
cn	廣告活動名稱	Str	<b>acquisition</b> : Google Ads gclid參數活動名稱 <b>Birthday_Gift</b> :生日禮 <b>eCoupon_Expiring</b> :即將到期優惠券

# 以迴圈合併多個檔案

檔名有一定規則，可用迴圈處理

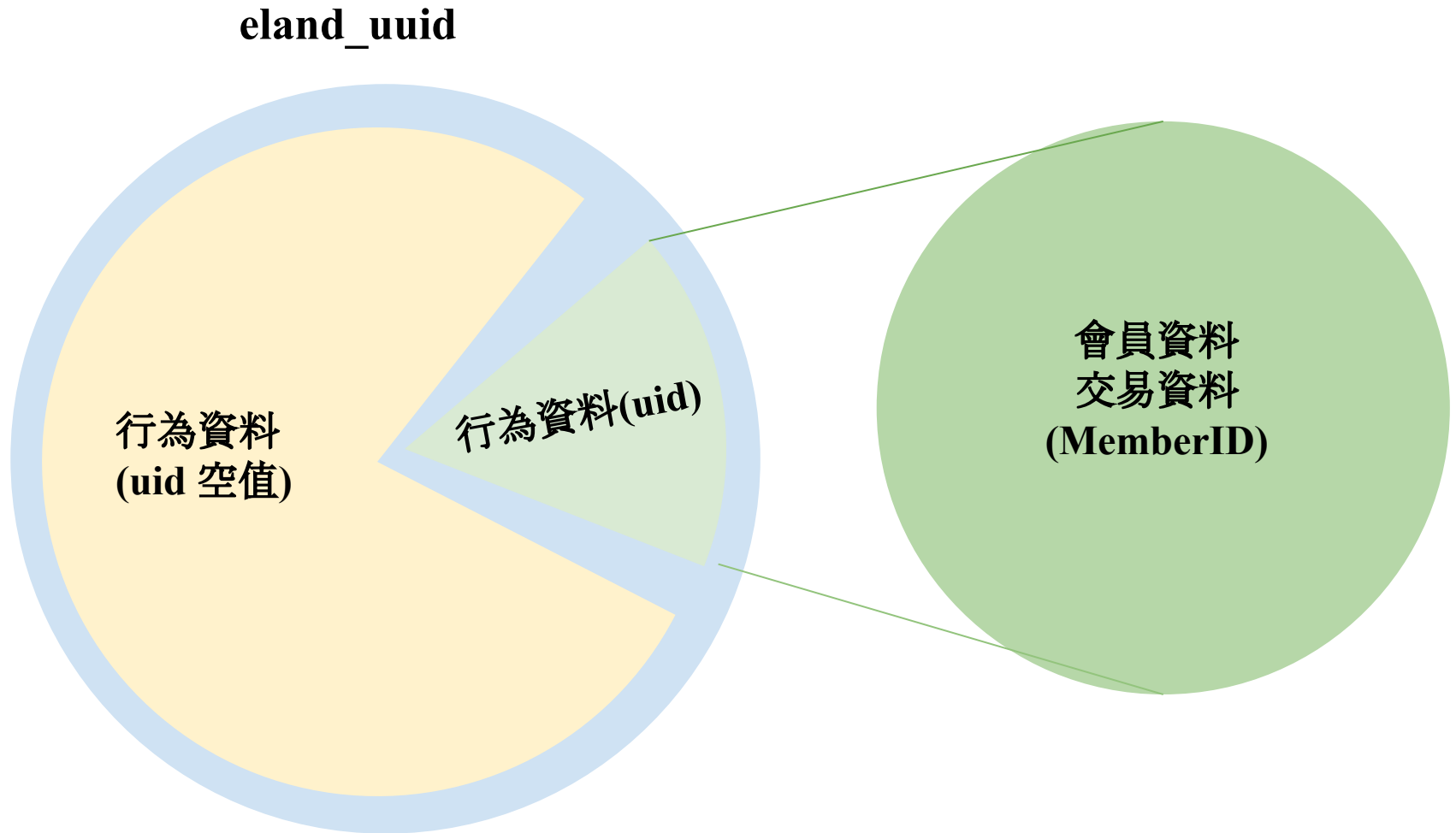
 activityPageView_2018-06-28	CSV 檔案
 activityPageView_2018-06-27	CSV 檔案
 activityPageView_2018-06-26	CSV 檔案

先讀取第一個檔案作為框架

```
1 df = pd.read_csv('C:/Users/USER/Downloads/activityPageView_2018-06-26.csv')
2 for i in [27, 28]: 再用迴圈將其他檔案合併上去(必須欄位相同)
3     path = 'C:/Users/USER/Downloads/activityPageView_2018-06-' + str(i) + '.csv'
4     df1 = pd.read_csv(path)
5     df = pd.concat([df, df1], axis=0)
6
7 print(pd.read_csv('C:/Users/USER/Downloads/activityPageView_2018-06-26.csv').shape)
8 print(pd.read_csv('C:/Users/USER/Downloads/activityPageView_2018-06-27.csv').shape)
9 print(pd.read_csv('C:/Users/USER/Downloads/activityPageView_2018-06-28.csv').shape)
10 print(df.shape)
```

```
(882, 9)
(1021, 9)
(939, 9)
(2842, 9)
```

# 資料集互相對應關係



# 資料集分析建議

- 同天同會員訂單合併成一張

```
order.groupby(['交易會員','交易日期']))['交易金額'].transform(sum)
```

- 訂單狀態條件

計算淨營收 / 淨銷售 -> 列入 Finish Return

計算所有交易行為 -> 排除Fail

計算營收 -> 排除Return Cancel

## Python 資料分析常用套件及語法

numpy  
pandas  
matplotlib

Group by 統一計算

merge 數據結合

concat 數據結合

isin / ~isin 是否(無)在指定範圍內

drop\_duplicates 去除重複

sort\_values 排序



# 資料集應用與方向探討

**OrderData**  
計算每個人交易次數

**MemberData**  
註冊日期配合交易

購物2次以上

購物一次

ACTIVE

POTENTIAL

NEW

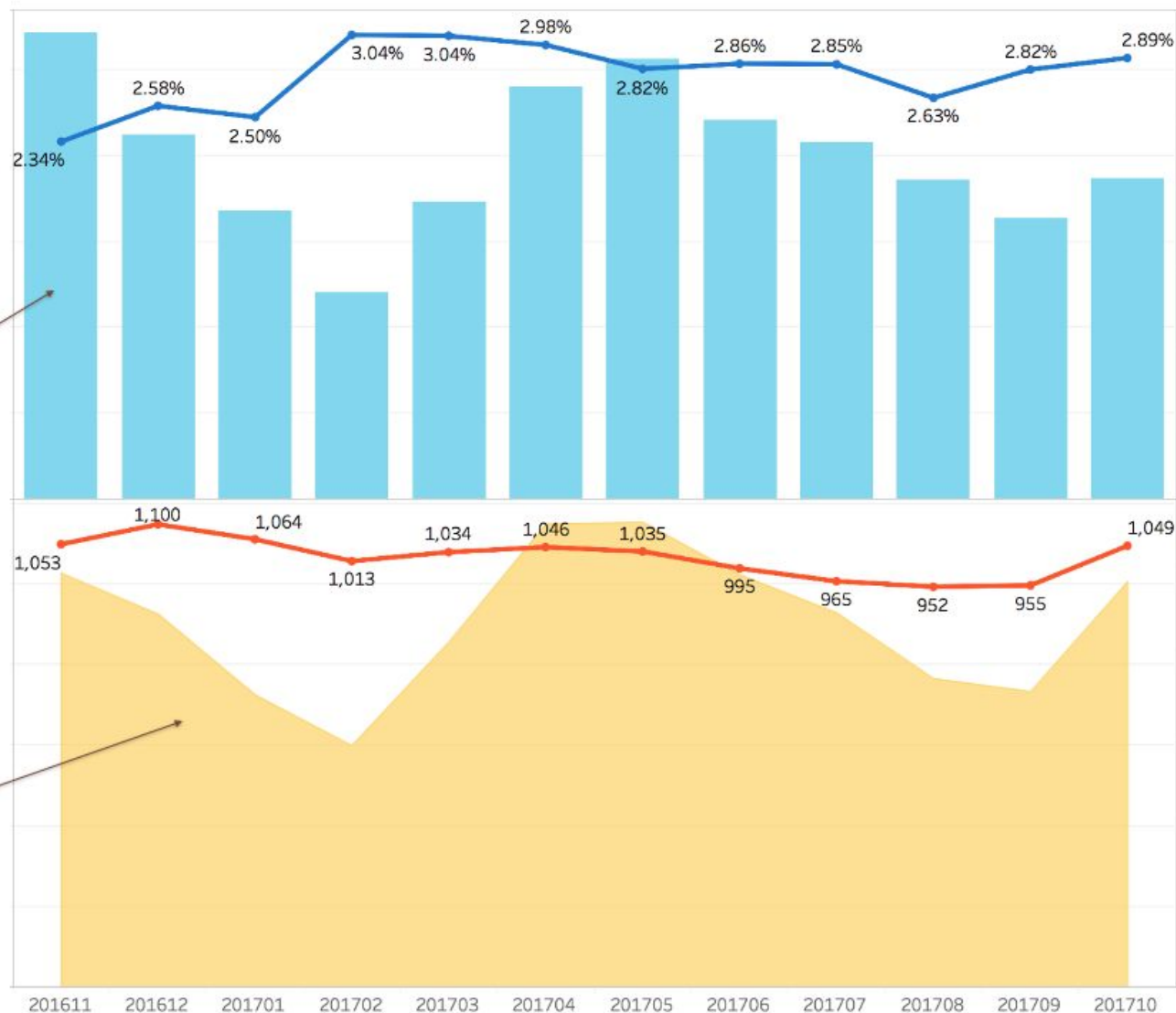
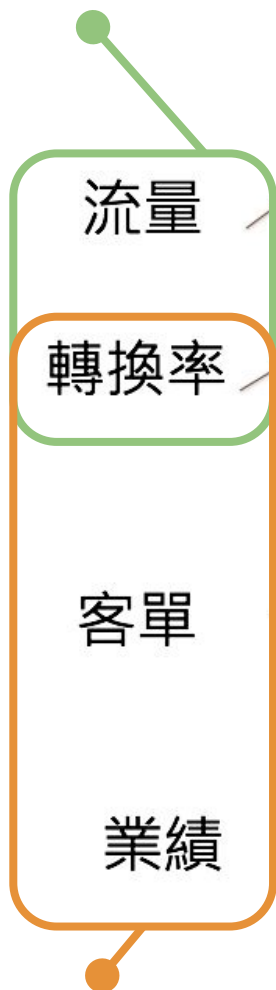
LOST

3X

購物週期倍數

**OrderData**  
計算購物天數中位數

## BehaviorData 瀏覽訪客數



## OrderData 交易金額 交易人數

用戶活躍度指標

OrderData

交易日期

交易金額

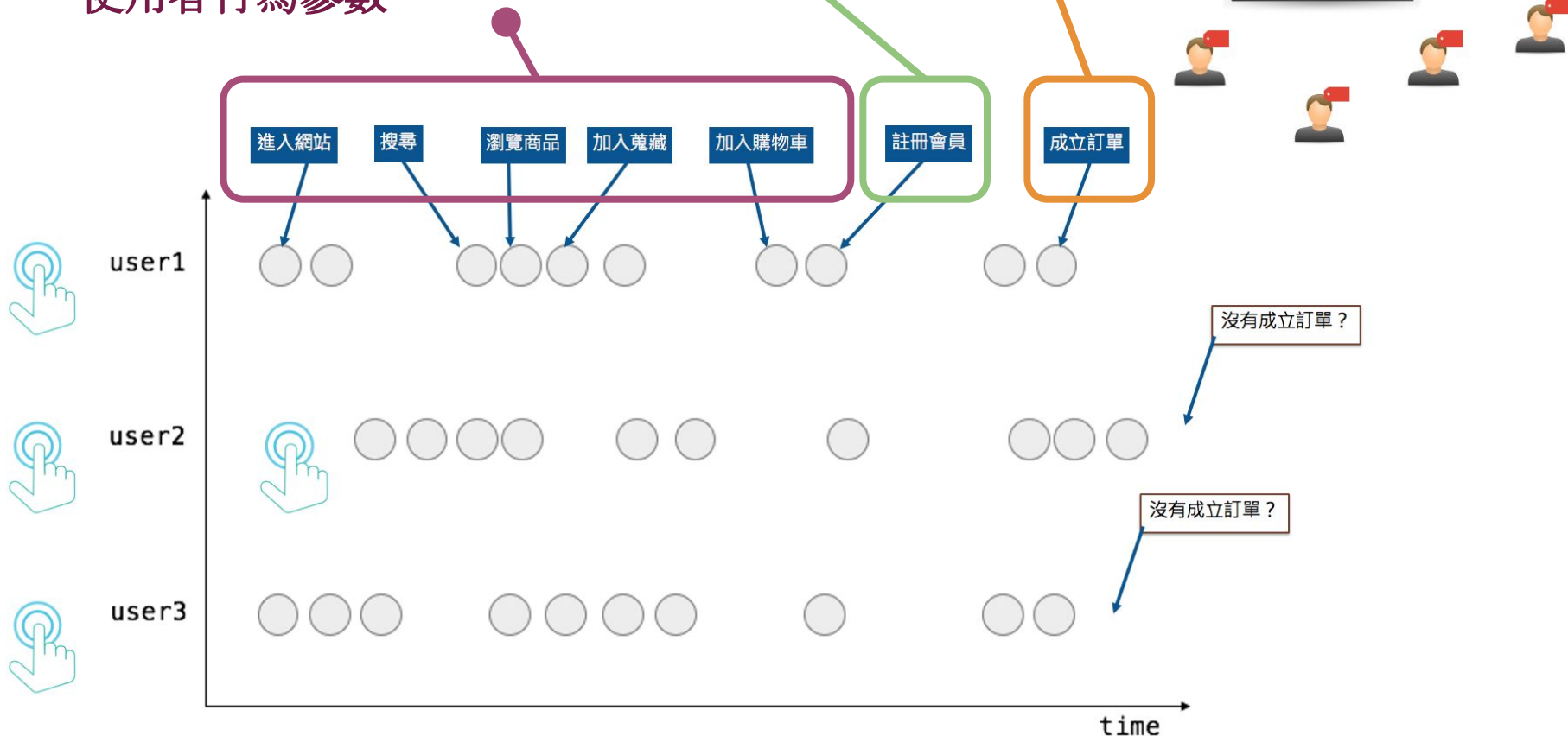
MemberData

會員註冊日期

BehaviorData

點擊日期

使用者行為參數



# 方向探討

**會員輪廓**

**消費行為**

**精準分群**

**行銷分析**

**導流操作**

**流失客群**

**趨勢預測**

**通路比較**

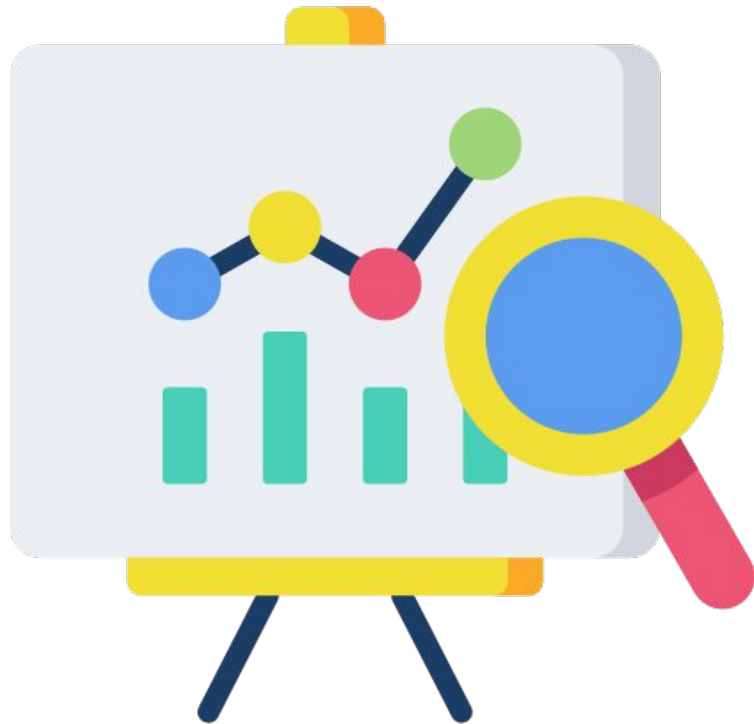
# 方向探討

MemberData
IsApplInstalled
IsEnableEmail
IsEnablePushNotification
IsEnableShortMessage
MemberCardLevel

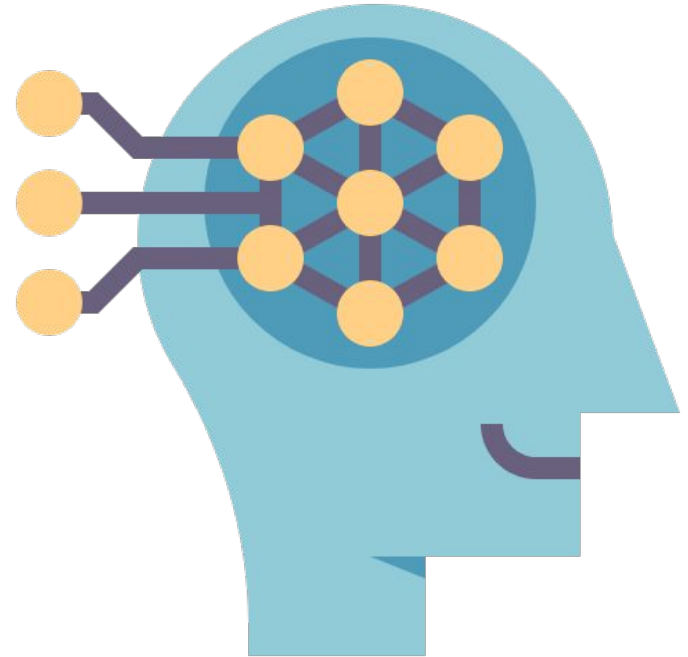
OrderData / OrderSalveData
ChannelDetail
PaymentType
ShippingType
TsCount
Qty
TotalPromotionDiscount
TotalCouponDiscount
TotalLoyaltyDiscount
OuterProductSkuCode
ProductSkuCode
SalePageCode

Behavior
pr1id
pr1pr
pr1qt
cu
tr
ti
cs
cm
cn

# 方向探討



敘述統計  
P-Value檢定  
r square  
常態分佈  
統計顯著性  
卡方分布



決策樹  
分群分類  
關聯分析  
隨機森林  
類神經網路  
集成模型

[WWW.91app.com](http://WWW.91app.com)

**91APP**  
品牌新零售專家

91App 產品處數據服務組  
CocoChung  
Email : [cocochung@91app.com](mailto:cocochung@91app.com)