

化妝品電商數據分析

使用工具：Python, Tableau

專案Python Code 連結：[GitHub](#)

Tableau Public 連結：[總覽](#)、[消費者分群分析](#)、[產品銷貨分析](#)

許為訢 Arthur Hsu



arthur752102@gmail.com



<https://github.com/endlessnoc>



<https://www.linkedin.com/in/weihsin-hsu/>



Contents



1. 分析結果摘要

針對[化妝品電商數據資料集](#)進行分析，在傳統的RFM模型上加入新維度S(Seniority)，利用K-Means分群將顧客分成五群，並探討各客群概況與擬定經營建議。

	客群	人數	集群特性	經營機會	經營建議簡評
	一般顧客	33,762	<ul style="list-style-type: none">各方面表現居中消費行為與銷售表現快速下滑	次要經營機會點	增強回購率，避免轉變為沉睡顧客。一般顧客是所有客群的基石，須維持一定數量
	新進顧客	25,020	<ul style="list-style-type: none">近期才有消費紀錄其餘方面表現居中	優先經營機會點	增加客戶黏著度，力求讓此客群轉變為忠誠顧客，將可有效提升營運表現
	忠誠顧客	5,253	<ul style="list-style-type: none">消費頻率非常高，但消費單價低人均累計消費額位居第一且遙遙領先*轉化率非常高	次要經營機會點	人均累計消費額已經非常高，考量用品需求有其上限，因此盡可能維持此客群的消費頻率即可
	高購買力顧客	2,221	<ul style="list-style-type: none">消費單價非常高，但消費頻率低人均累計消費額次高明顯品牌偏好	優先經營機會點	增加客戶黏著度，在高消費單價的基礎之上增強消費頻率，借乘數效應放大銷售額
	沉睡顧客	26,396	<ul style="list-style-type: none">很久以前有來消費，但很久沒來了消費頻率不低，仍具經營價值	優先經營機會點	發送推播通知與折價券，喚醒此客群並盡可能增強回購率，使此客群維持消費習慣

註*：轉化率表示瀏覽商品之後有購買的比率

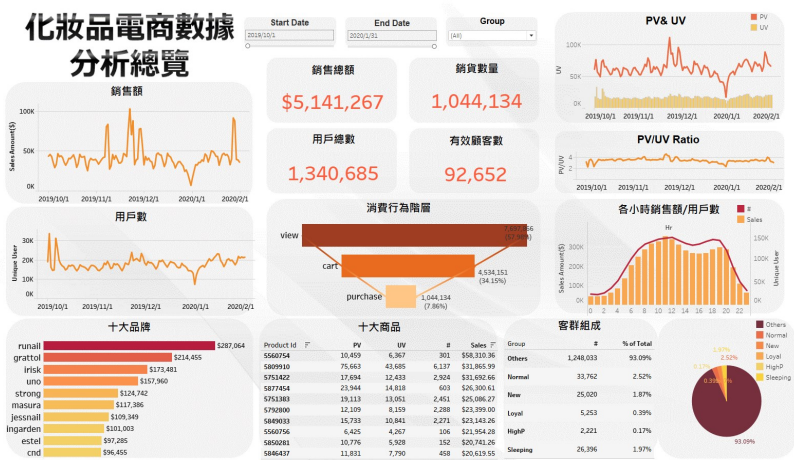
2. Tableau Dashboard

使用Tableau建構了三個幫助探索資料的Dashboard，分別為總覽、消費者分群分析、產品銷貨分析。可以選擇期間、客群來查看對應的各種數據，幫助瞭解趨勢以及客群之間的差異。

Tableau Public 連結：[總覽](#)、[消費者分群分析](#)、[產品銷貨分析](#)

總覽

總體經營狀況
e.g. 銷貨總額、用戶總數、銷貨數量



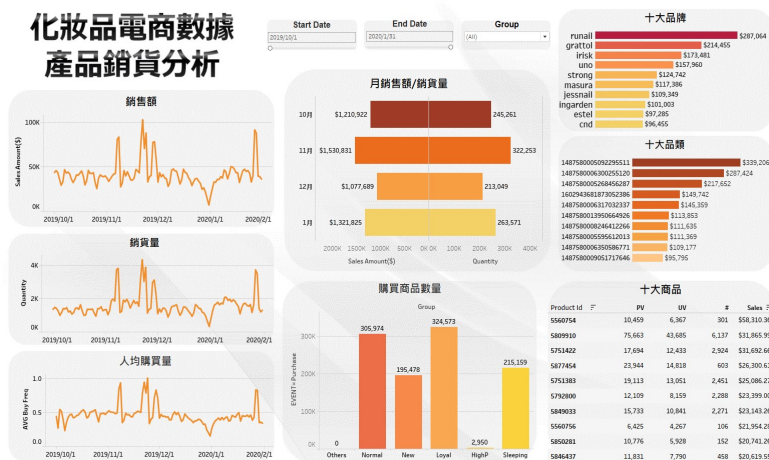
消費者分群分析

各種客群的銷售與消費行為
e.g. 各種客群의 PV, UV, 人均消費額



產品銷貨分析

品牌、品類、產品與銷貨數量
e.g. 每日銷貨數量、十大品牌/品類



3. 分析方法與流程

使用工具：Python(Pandas, Numpy, Matplotlib, Seaborn, SKlearn)
Tableau Public

目的：針對**化妝品電商數據資料集**進行分析，將用戶分群與進行分群研究，找出各組用戶的消費特性與發掘其中的商業價值，並且給予經營建議。

選定數據

- 使用前80%的數據(前四個月)做分析
- 最後使用完整期間數據(五個月)重複同樣步驟
- 將兩組集群結果對比，觀察在加入了最後一個月的數據後產生了什麼變化

RSFM模型分析

在傳統的RFM模型上加入新維度S，利用顧客的四種特性分析：

- R(Recency)最近一次消費
- S(Seniority)最久一次消費
- F(Frequency) 消費頻率
- M(Monetary)消費均價

K-Means集群

利用K-Means幫助集群分析，將具有共同特性的消費者分於同群，最後得到數群具有各自特色的客群

1

資料預處理

2

探索性資料
分析

3

集群分析

4

Tableau
Dashboard

5

結果分析

6

集群比對

7

結論與經營
建議

3-1. 資料預處理

1

去除重複值

觀察到有出現不合理的重複值，*保守起見先將所有重複值刪除

2

去除異常值

價格出現明顯不合邏輯的負值，直接將其刪除

3

處理NaN

- 品類代號的NaN比例達98%，不過有功能相同的品類ID可以使用，因此刪除品類代號欄位
- 品牌約42%的數據為NaN，都先替換成Unknown

4

新增時間欄

將各列時間拆解得出「月份」、「星期幾」、「小時」等欄位

```
0 event_time
1 event_type
2 product_id
3 category_id
4 brand
5 price
6 user_id
7 date
8 month
9 day
10 weekday
11 hr
```

註*：數據中出現多筆「瀏覽」行為的重複數據，那代表該使用者「同時瀏覽了一件商品2次」，這顯然不合邏輯。因為根據現有資料不確定是否可對同一件商品進行批量操作(加入/移出購物車、購買)，但既然瀏覽出了不合邏輯的狀況，那這邊都先將重複值視為異常刪除，並且保留此問題。若後續有機會確認此問題再來調整即可。

3-2. 探索性資料分析

從各個面向探索數據，主要發現了以下狀況：



時間面

- 從月份級別而言，**11月**是營運高峰期，**12月**表現則最疲弱
- 從日內級別而言，四種消費者行為(瀏覽, 添加&移除購物車, 購買)數量的高峰期位於每天的**12點**和**19點**，這兩個時段的銷售額表現亦最突出
- 營運表現在假日並沒有比較突出



產品面

- 銷售額占比最高的商品超過了**1%**，剩餘商品的銷售額占比曲線頗為平滑
- 前十大品牌(排除Unknown後)銷售額佔比約三成、前十大品類占比約三成，整體曲線亦都相當平滑
- 不論是品類、品牌、商品都頗符合二八法則，即前**20%**的項目占了約**80%**的銷售額
- 綜上述，在產品方面並無一家獨大的壟斷狀況，產品銷售結構健康



顧客面

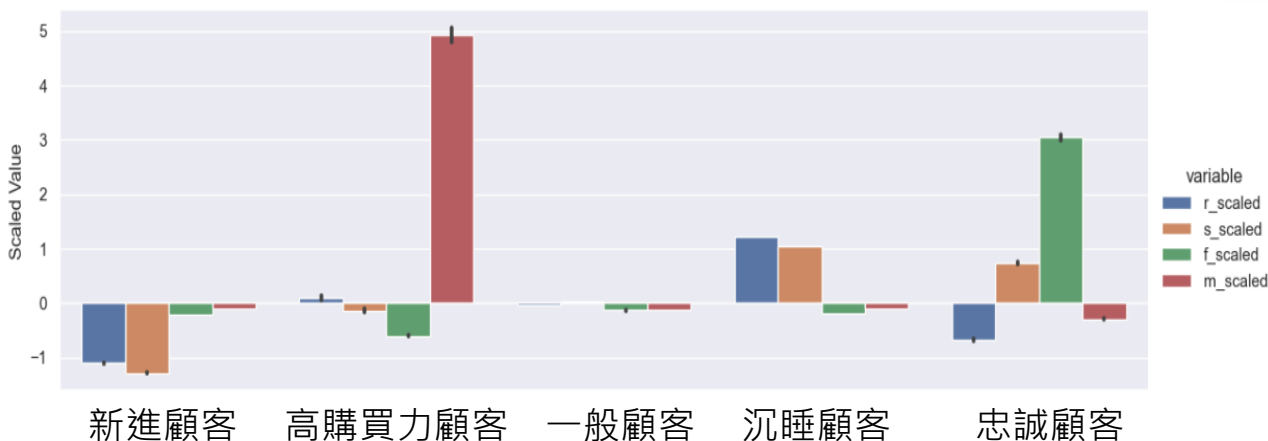
- 有消費紀錄的消費者比例僅佔數據中顧客約**7%**
- 轉換率 (瀏覽後有購買的比率) 低於**10%**
- 顧客累計消費額的曲線平滑，沒有超級大戶的存在 (第一名的顧客消費額僅佔總消費額**0.06%**)

3-3. 顧客集群分析

將資料根據顧客ID彙總後，得出有消費紀錄的顧客共有92,652位，再計算每位顧客的：

- r (recency): 最近一次消費距*今天的日期數，越低越好，代表顧客仍然有持續來消費的習慣
- s (seniority): 最久一次消費距今天的日期數，沒有絕對好壞，越高代表顧客很久以前就有來消費過，越低則代表是較新的客戶
- f (frequency): 消費次數，越高越好，代表顧客來消費的次數越多
- m (monetary): 平均消費單價，越高越好，代表顧客每次的消費額越高

由於數據單位不一，因此完成數據標準化後使用K-Means來做集群分析。（經Elbow Method後選擇分成5群）



△數值標準化後的各客群r, s, f, m比較，
可以觀察到各客群的特點

group	count	r_range	s_range	f_range	m_range
新進	25020	3.0-35.0	3.0-37.0	1.0-18.0	2.33-13.7
高購買力	2221	11.0-108.0	12.0-108.0	1.0-2.0	44.29-122.22
一般	33762	31.0-74.0	45.0-86.0	2.0-21.0	2.25-12.93
沉睡	26396	82.0-118.0	84.0-119.0	2.0-18.0	2.28-13.56
忠誠	5253	3.0-73.0	57.0-119.0	36.0-98.0	2.31-6.47

△各客群r,s,f,m 的人數與middle 80%(10th ~ 90th)數據，
可以清楚看出各客群在各維度上的具體差距

*註：分析時將數據中的最晚的日期+1天當作「今天」

3-3. 顧客集群分析

經由 r , s , f , m 集群分析分出的五群，稍加整理之後可以得出各群特色：



一般顧客

在各方面都居中，跟其他四群相比沒有明顯特色



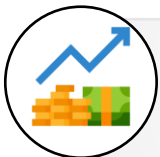
新進顧客

r 和 s 都較低，最近才有消費紀錄，因此距上次消費的時間也較近



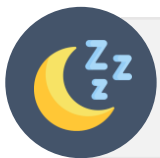
忠誠顧客

f 非常高但 m 偏低，消費頻率非常高，但是消費單價較低



高購買力顧客

m 非常高但 f 偏低，消費單價非常高，但是消費頻率較低

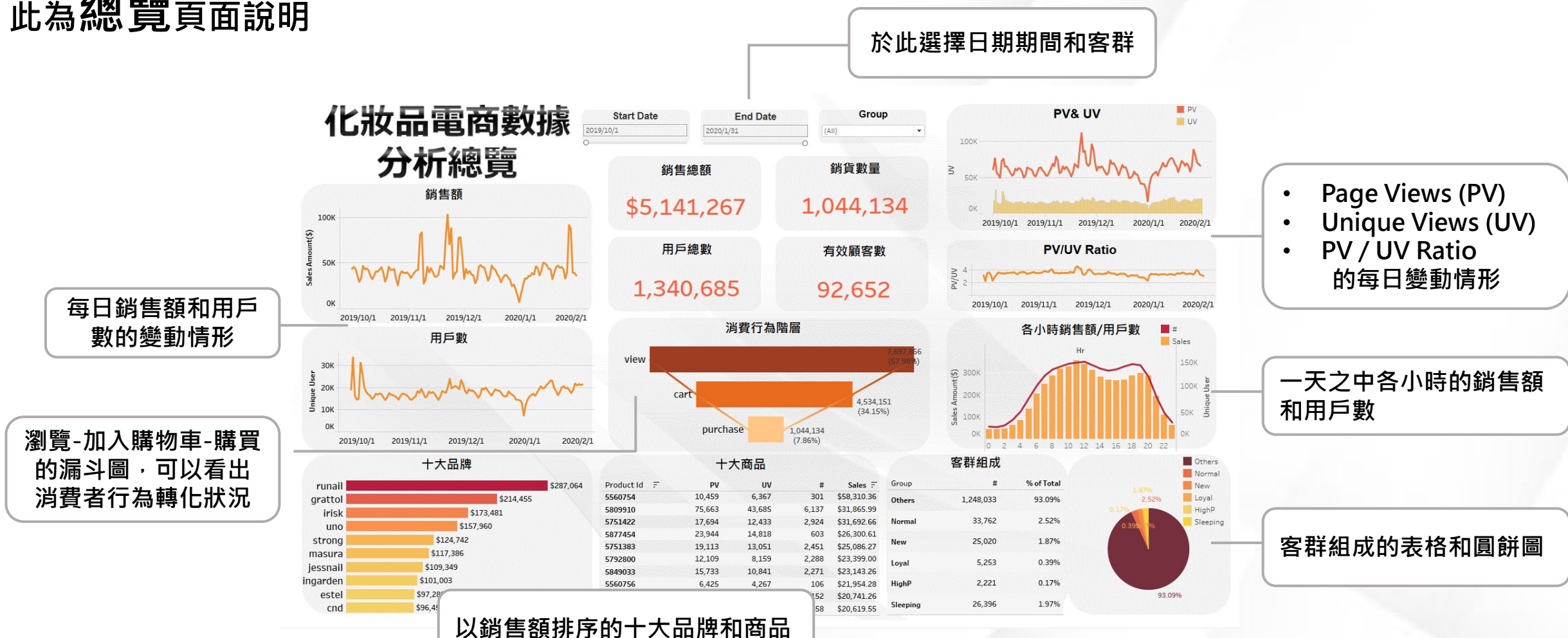


沉睡顧客

r 與 s 都較高，很久之前就有消費紀錄，但也一陣子沒有來消費了

3-4. 製作Tableau Dashboard

使用Tableau建構了三個幫助探索資料的Dashboard，分別為總覽、消費者分群分析、產品銷貨分析。
此為**總覽**頁面說明



註*：有效顧客意指「有消費紀錄」的顧客
註*：品牌已排除Unknown

3-4. 製作Tableau Dashboard

使用Tableau建構了三個幫助探索資料的Dashboard，分別為總覽、消費者分群分析、產品銷貨分析。
此為消費者分群分析頁面說明

化妝品電商數據 消費者分群分析

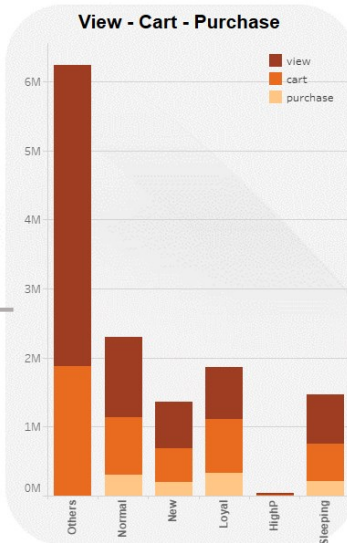
於此選擇日期期間和客群

Start Date
2019/10/1

End Date
2020/1/31

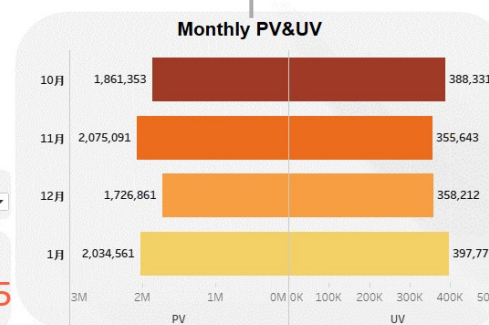
Group
(All)

用戶總數
1,340,685

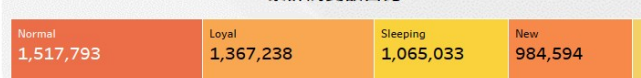


各客群
瀏覽-加入購物車-購買
的累積柱狀圖，可以看
出各種客群的行為轉化

每個月累計的UV和PV



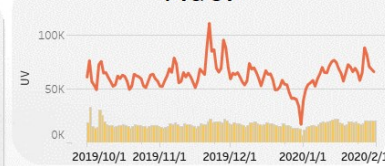
累計消費額占比



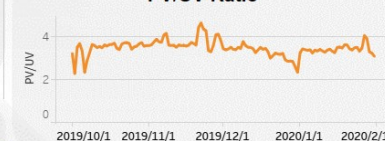
Group	#	PV	UV	PV/UV	Avg Contribution	Avg Freq	# of Purchase	View - Purchase %	Sales
Others	1,248,033	4,371,589	1,218,314	3.50	\$0	0.000	0	0.00%	
Normal	33,762	1,166,088	32,203	34.54	\$45	9.063	305,974	26.24%	\$1,517,793
New	25,020	669,401	23,571	26.75	\$39	7.813	195,478	29.20%	\$984,594
Loyal	5,253	752,220	5,204	143.20	\$260	61.788	324,573	43.15%	\$1,367,238
HighP	2,221	24,594	2,128	11.07	\$93	1.328	2,950	11.99%	\$206,609
Sleeping	26,396	713,974	24,895	27.05	\$40	8.151	215,159	30.14%	\$1,065,033

各客群詳細數據表格，
包含人數，PV, UV, PV/UV, 平均消費
次數、總消費額.....等等

PV& UV

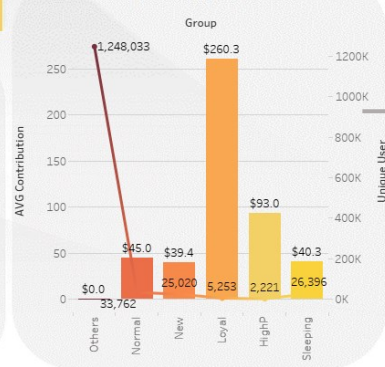


PV/UV Ratio



- Page Views (PV)
- Unique Views (UV)
- PV / UV Ratio
的每日變動情形

人均累計消費額&人數



各客群的人均累計消費額
與累計人數圖表

3-4. 製作Tableau Dashboard

使用Tableau建構了三個幫助探索資料的Dashboard，分別為總覽、消費者分群分析、產品銷貨分析。
此為產品銷貨分析頁面說明

於此選擇日期期間和客群

化妝品電商數據 產品銷貨分析

Start Date
2019/10/1

End Date
2020/1/31

Group
(All)

十大品牌



十大品類



十大商品

Product Id	PV	UV	#	Sales
5560754	10,459	6,367	301	\$58,310.36
5809910	75,663	43,685	6,137	\$31,865.99
5751422	17,694	12,433	2,924	\$31,692.66
5877454	23,944	14,818	603	\$26,300.61
5751383	19,113	13,051	2,451	\$25,086.27
5792800	12,109	8,159	2,288	\$23,399.00
5849033	15,733	10,841	2,271	\$23,143.26
5560756	6,425	4,267	106	\$21,954.28
5850281	10,776	5,928	152	\$20,741.26
5846437	11,831	7,790	458	\$20,619.55

以銷售額排序的10大

- 品牌
- 品類
- 商品

每日銷售額變動



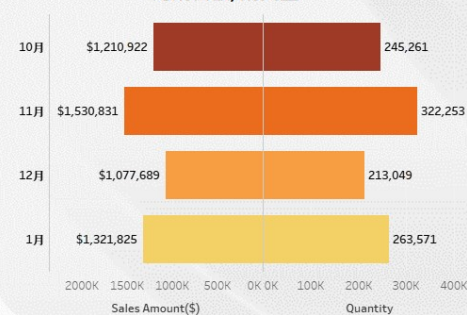
每日銷貨量變動



每日人均購買量



月銷售額/銷貨量



購買商品數量



各客群累計購買商品數量

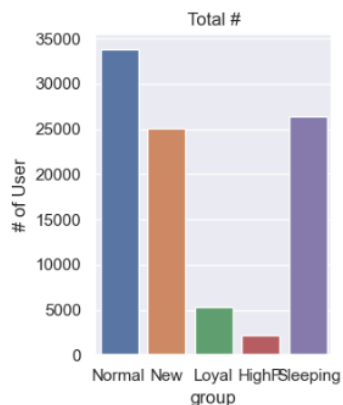
註*：品牌已排除Unknown

3-5. 集群結果分析

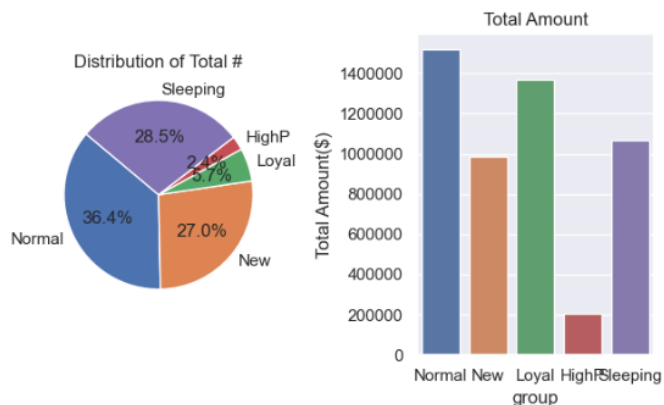
藉由Python查看各群特性的同時，也根據數據確保分群結果無誤，最後統計出此表：

group	累計消費額	人數	平均累計消費額	平均消費單價	平均消費次數
一般	1,517,793	33,762	45	4.96	9.06
新進	984,594	25,020	39.4	5.03	7.81
忠誠	1,367,238	5,253	260.3	4.21	61.79
高購買力	206,609	2,221	93	70.03	1.33
沉睡	1,065,033	26,396	40.3	4.95	8.15

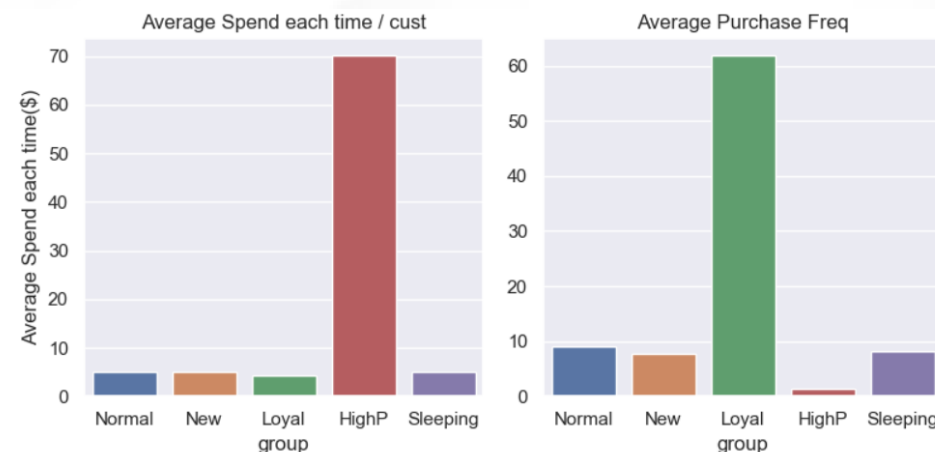
圖表節錄：



△各客群人數長條圖與圓餅圖



△各客群總消費額長條圖與圓餅圖



△各客群人均每次消費單價

△各客群人均消費頻率

3-5. 集群結果分析

從上頁由Python統計出的各群特性表格輔以Tableau的視覺化呈現，觀察到了以下狀況：



一般顧客

r, s, f, m四方面都居中。但此客群的表現正面臨快速衰退的窘境，四種消費者行為從11月巔峰後便呈現下滑趨勢，銷售額也有頗為嚴重的衰退幅度，平均活躍度下降，需要謹慎提防變成沉睡客戶。



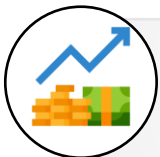
新進顧客

r和s都較低。這個客群在各方面都和一般顧客頗為相似，可視作是大部分一般顧客的前身，營運重點可以聚焦於如何使這群顧客轉變為忠誠顧客或高購買力顧客。



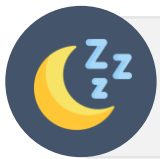
忠誠顧客

f非常高但m偏低。雖然客群人數並不多，但人均累計消費額非常高，因此也佔有頗大比例的銷售額。在轉化率方面表現優秀，是非常重要的一群客戶。



高購買力顧客

m非常高但f偏低。雖然客群人數並不多，但消費均價相當高，在人均累計消費額位居第二，此客群有非常明顯的品牌偏好(strong)。多數顧客在四個月期間僅有1或2次消費紀錄、且PV/UV比率與轉化率都敬陪末座，如何提升黏著度是關鍵。



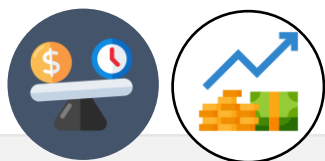
沉睡顧客

r和s都較高。此客群整體其實消費頻率和轉化率都不低，但表現在11/9迎來斷崖式下跌，銷售額在一周後只剩原先的5%，是否發生什麼重大事件導致消費者出走？（其實全體數據在11月過後都是持續下降，推測可能是活動結束後沒有適當的營運）

3-6. 集群對比

使用全部五個月的數據重複一次相同步驟的集群分析，對比兩組集群結果，觀察同樣的消費者在加入了最後一個月的數據後，分群結果產生了什麼變化。

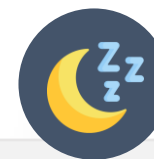
原先集群分析的92652位顧客，有12195位 (11%) 的集群發生了變化。絕大多數變化都是：



一般、高購買力顧客

最容易變成.....

- 一般顧客的變化可以從前述Tableau分析中窺見一二
- 高購買力顧客因為消費頻率低，若在二月也沒有消費就容易被歸類成沉睡顧客



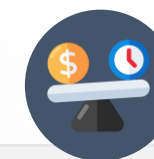
沉睡顧客



新進、忠誠、沉睡顧客

最容易變成.....

- 合乎邏輯的變化
- 大數法則下的均值回歸？
- 如何使忠誠顧客轉變為一般顧客比例下降、沉睡顧客轉變為一般顧客比例上升？



一般顧客

3-7. 結論與經營建議

分析小結：這家化妝品電商的产品銷售結構穩定，不管是品牌、品類、產品之間的表現都頗為均衡。但消費者行為變化較大導致銷售表現也有所起伏，在客戶營運方面尚待加強。

分群分析給予各客群經營建議：



一般顧客

一般顧客是人數最多的客群，也是各種客群轉變的重要中介站。維持一定數量的一般顧客對於顧客結構會有相當程度的幫助。目前面臨一月消費行為和營運表現急遽衰退的危機，需要提升回購意願，可以從發放折扣、滿額折價等面向規劃。



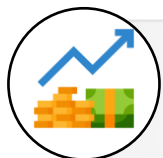
新進顧客

由於最早消費紀錄在不久之前，因此會是可塑性最強的客群。經營目標應盡可能加強其黏著度，針對推送推薦商品投放、活躍度系統等方面加強，目標訂於將此客群轉為忠誠顧客（尤佳）或者高購買力顧客。



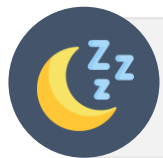
忠誠顧客

消費頻率高，連帶使人均累計消費額也非常高。考量到化妝品在需求面有其上限，且高消費頻率很可能使此客群的偏好定型，因此可主要以維持其消費頻率為目標即可。



高購買力顧客

有很明顯的品牌偏好(strong)但轉化率較低，非常有可能有品牌迷思，可以和品牌方合作針對此類客群推出專屬活動。此客群的消費單價非常高，只要能增加消費頻率，便可以此為基礎透過乘數效應放大數字，大幅增加人均消費額。



沉睡顧客

絕大多數的消費者行為止於11月中旬，可以先探清緣由並以喚醒客群為目標做行銷規劃。人數位居第二，仍然具備經營價值，若是能吸引他們回來消費將可以提升大量的用戶基數，可以發送推播通知並且以邀請碼活動吸引此客群回歸。

End

**Thank you for your
kind attention.**

Icon Attribution in this Project:

[Free icons designed by iconixar / Flaticon](#)

[Free icons designed by ultimatearm / Flaticon](#)

[Freepik / Create great designs, faster](#)



arthur752102@gmail.com



<https://github.com/endlessnoc>



<https://www.linkedin.com/in/weihsin-hsu/>