91app商業數據分析——顧客分群與預測購買者行為預測

1. 資料處理——整理feature

根據91App給的原始資料，分別是BehaviorDataForNTU, MemberDataForNTU, OrderDataForNTU, OrderSlaveDataForNTU做資料清洗後，清洗出以下features的檔案，存成91App\_features.txt。

* Session\_number：總session次數。
* Off\_return\_item\_number：線下退換貨總商品數。
* Off\_return\_frequency：線下總退換數次數。
* Buy\_time\_without\_offline\_return：排除線下退換後的總購買次數(即total\_F 減去 off\_return\_frequency)。
* Online ratio：線上購買佔所有購買金額(線上+線下)的比例。
* Total day using percentage：一段時間中使用過線上商店的天數百分比。
* View per date：平均一天使用次數。
* Total conversion rate：買過的商品種類 / 瀏覽過的商品種類。
* Off\_cart\_c：線下買的商品種類 / 放在購物車的商品種類。
* Off\_fav\_c：線下買的商品種類 / 放在我的最愛的商品種類。
* Off\_view\_c：線下買的商品種類 / 瀏覽過的商品種類。
* On\_cart\_c：線上買的商品種類 / 放在購物車的商品種類。
* On\_fav\_c：線上買的商品種類 / 放在我的最愛的商品種類。
* On\_view\_c：線上買的商品種類 / 瀏覽過的商品種類。
* Off\_mix\_c：線下買的商品種類 / 加入購物車或者cart的商品種類。
* On\_mix\_c：線上買的商品種類 / 加入購物車或者cart的商品種類。
* Total\_mix\_c：線上、線下買的商品種類 / 加入購物車或者cart的商品種類。
* Viewtime\_ave：每個人平均每次瀏覽頁面時間。
* Viewtime\_med：每個人每次瀏覽頁面時間的中位數。
* Cart\_med\_time：每個人平均放到購物車時間的中位數。
* Cart\_ave\_time：每個人平均放到購物車的時間(以秒計算)
* Cart\_within\_3：購物車內商品三小時內被購買的比率。
* Cart\_within\_24：購物車內商品二十四小時內被購買的比率。
* View\_count\_med：消費者購買A商品後，過去曾看了A商品頁面多少次的中位數。
* View\_count\_ave：消費者購買A商品後，過去曾看了A商品頁面多少次的平均數。
* Total\_F：十四個月總購買金額。
* Total\_M：十四個月總購買次數。
* Total\_S 總瀏覽行為次數
* F List：每個月購買次數。
* M list：每個月購買金額。
* Total discount percentage：總折扣金額佔總購買金額的比例。

1. 資料處理————分群方式

根據以上洗好的檔案，進行如下步驟：

第一階段分群

講所有資料根據消費習慣分為五群：

step1

依照total\_day\_using\_percentage <2%(約使用天數少於8天)  
小於2% => inactive

超過2% => 繼續過篩

step2

看有無買過東西

無 => df\_never\_buy，這些人可能以後會買，還要再預測

有 => 繼續過篩

step3

全部都在線下購買

是: df1，offline buyers，(根據online\_ratio\_without\_offline\_return 即除去有線下退換貨行為的購買者的線上消費比率<= 20%)劃分

否: df2，online buyers繼續過篩

step4

依照view conversion rate （線上瀏覽——購買的轉換率）分三等份:

小於33百分位 => 深思熟慮型df3  
大於67百分位 => 衝動購買型df4

介於33, 67 之間=> 一般型df5

第二階段分類

將以上5個群，依次實行以下過程，每一個群的每個客戶，根據每個月購買行為劃分(Ex. df1有10000個客戶，則該群每個人都有14個月的分類，每個月都有一個類別)：

step5

根據每個月是否有瀏覽紀錄劃分

無 => 0類客戶

有 => 繼續過篩

step6

將有瀏覽紀錄卻無購買紀錄的客戶分群，根據是否大於12劃分（12是根據探索性資料分析EDA得出）

view小於12且無購買行為 => 1類客戶

view大於12 且無購買行為 => 2類客戶

step7

將有購買紀錄的客戶分群，將客戶分成4類，放在四個象限，x 軸為購買頻率frequency，y軸為購物金額money，由EDA得出frequency的mean值=1，money的mean值=2280

M

4類客戶,

F<1,M>2280

3類客戶,

F>1,M>2280

F

5類客戶,

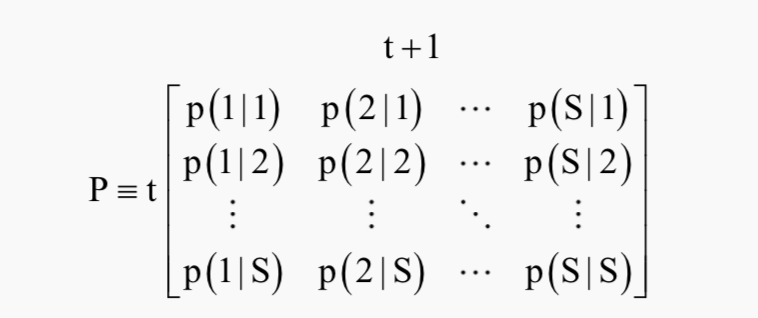
F<1,M<2280

6類客戶,

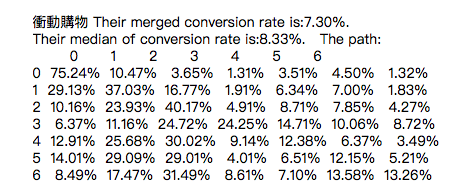
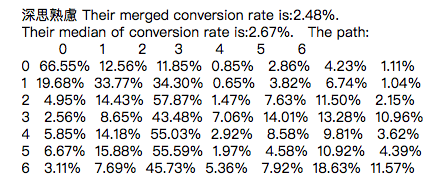
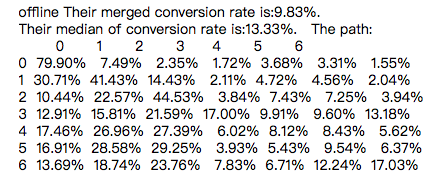
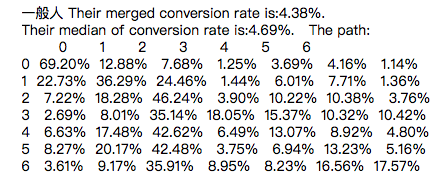
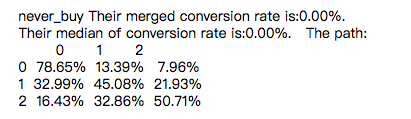
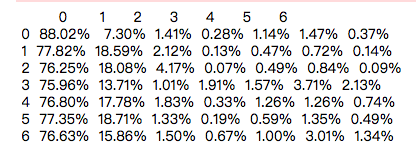
F>1,M<2280

1. 資料處理——群內消費者行為分析

利用馬可夫一階移轉矩陣，用過去十三個月(以群為單位)的消費者行為數據，判定消費行為屬於以上哪一象限及遷徙矩陣，找出每一階段的變化規律，如衝動購物型的人，從2型到3型的機率為40.17%



inactive客群



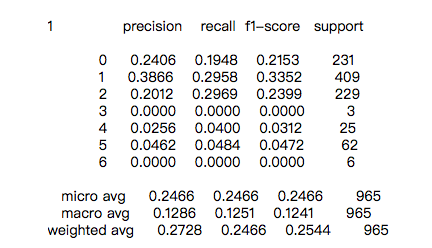
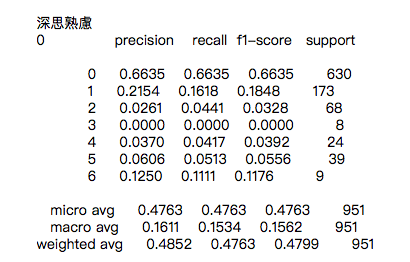
1. 群內消費者行為預測分析

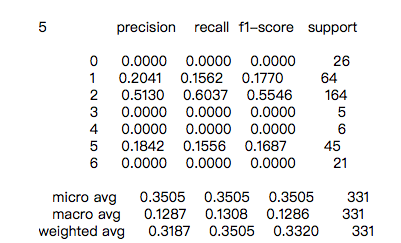
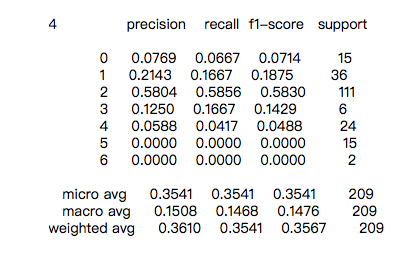
根據上述統計資料的規律，預測下一期(第14個月)消費者的遷徙狀態如下：

以深思熟慮型為例，深思慮型客戶從0型到0型的機率最高，為66.35%

從1型到1型的概率也最高，38.66%，從4型到2型的可能性最高，為58.04%

從5型到2型的可能性最高，為51.3%





1. 其他發現

加入購物車或者最愛後若高過一天，必定不買

