

電商顧客流失分析

 [程式碼連結](#)

Euan Hsia 夏震華



Contents

- 1 專案背景
- 2 探索性資料分析
- 3 模型建立
- 4 主要挑戰
- 5 未來應用

資料欄位介紹

欄位名稱	欄位介紹
CustomerID	客戶唯一識別碼
Churn	流失標記（是否流失）
Tenure	客戶在公司的服務時間
PreferredLoginDevice	客戶偏好的登入裝置
CityTier	城市等級（城市階層）
WarehouseToHome	倉庫到客戶住家的距離
Gender	客戶性別

欄位名稱	欄位介紹
HourSpendOnApp	客戶在網站或App上的平均使用時數
NumberOfDeviceRegistered	客戶註冊的裝置數量
PreferedOrderCat	客戶上個月偏好的訂購類別
SatisfactionScore	客戶對服務的滿意度評分
NumberOfAddress	客戶新增的地址數量
Complain	客戶上個月是否有提出投訴
OrderAmountHikeFromlastYear	訂單金額相較去年成長的百分比

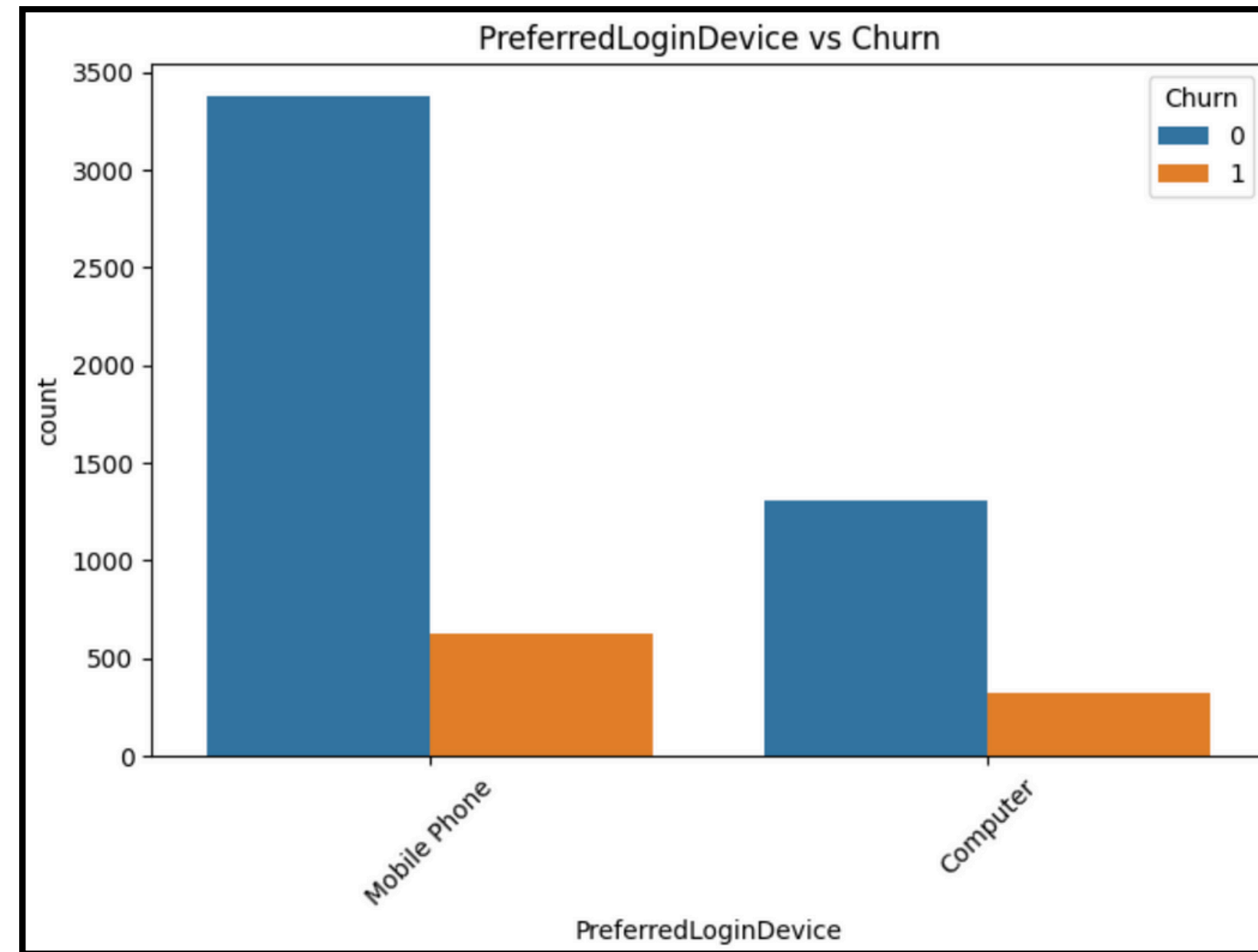
欄位名稱	欄位介紹
CouponUsed	上個月使用的優惠券數量
OrderCount	上個月下單的總次數
DaySinceLastOrder	距離客戶上次下單的天數
CashbackAmount	上個月獲得的平均現金回饋金額



目標

根據提供欄位預測顧客是否會流失

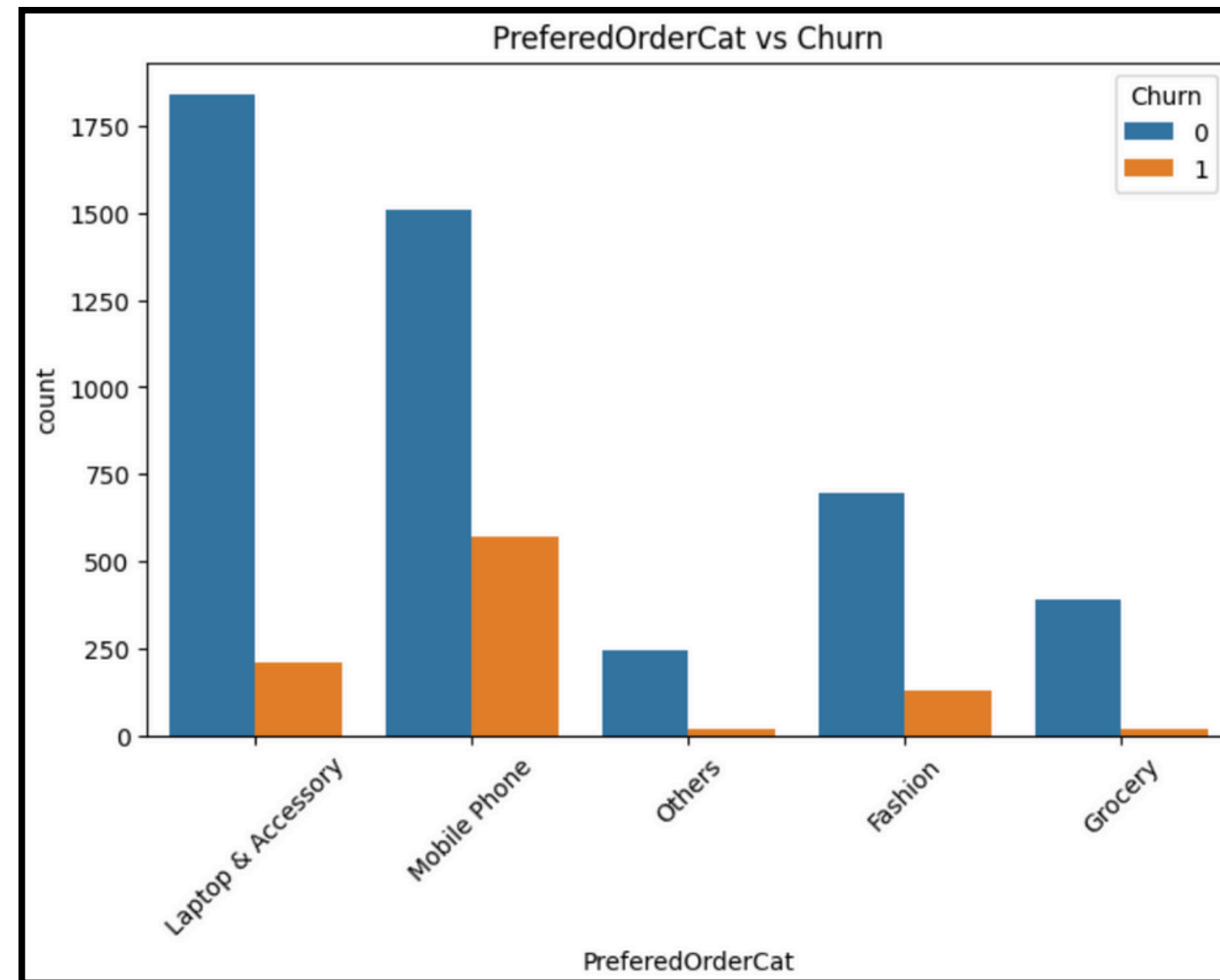
使用裝置分佈



分析

大多數用戶使用智慧型手機登入app

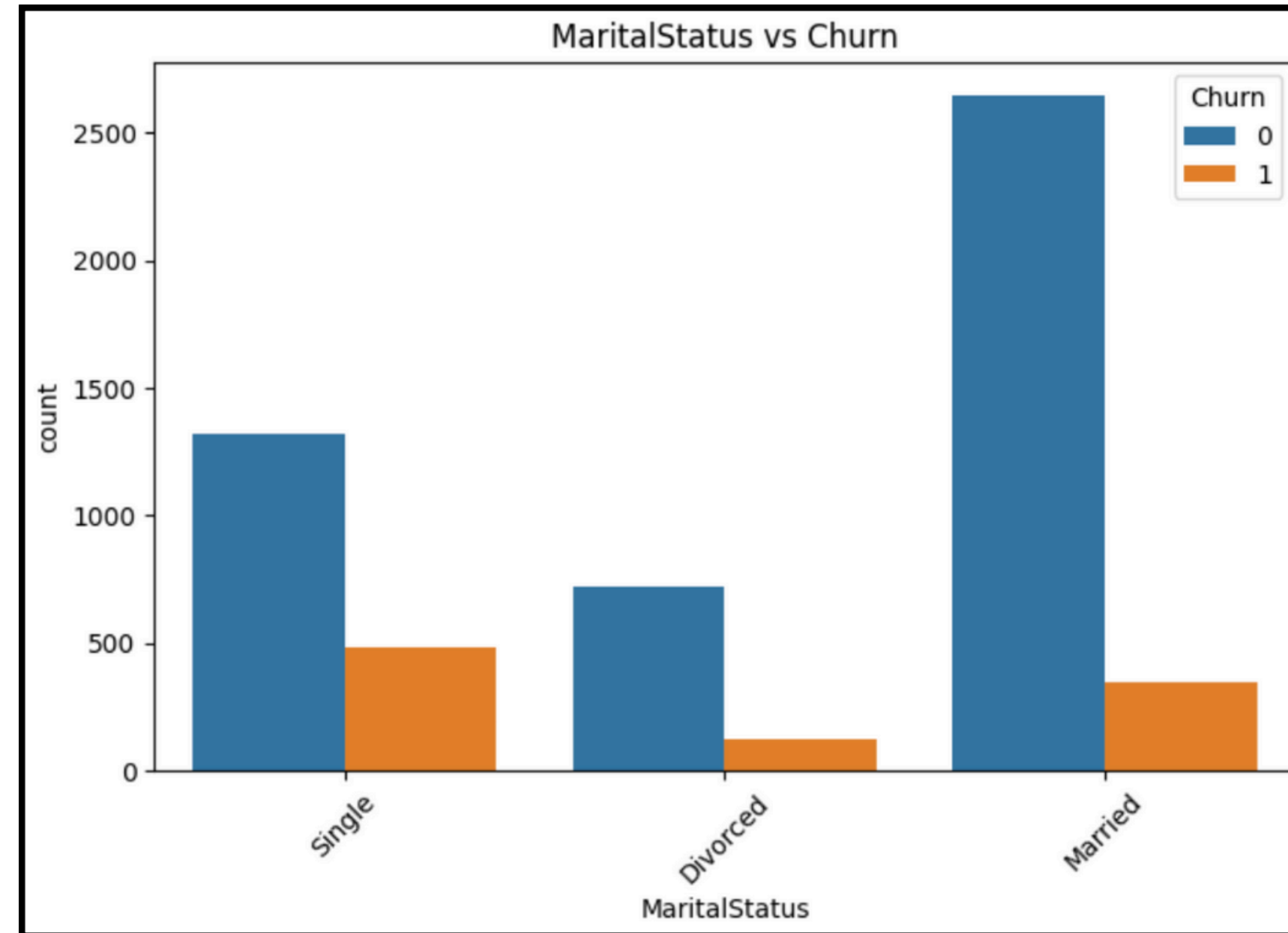
商品類別偏好



分析

筆電及周邊商品為最熱門類別，第二則是智慧型手機，其中我們可以發現智慧型手機的用戶流失率相較其他類別較高

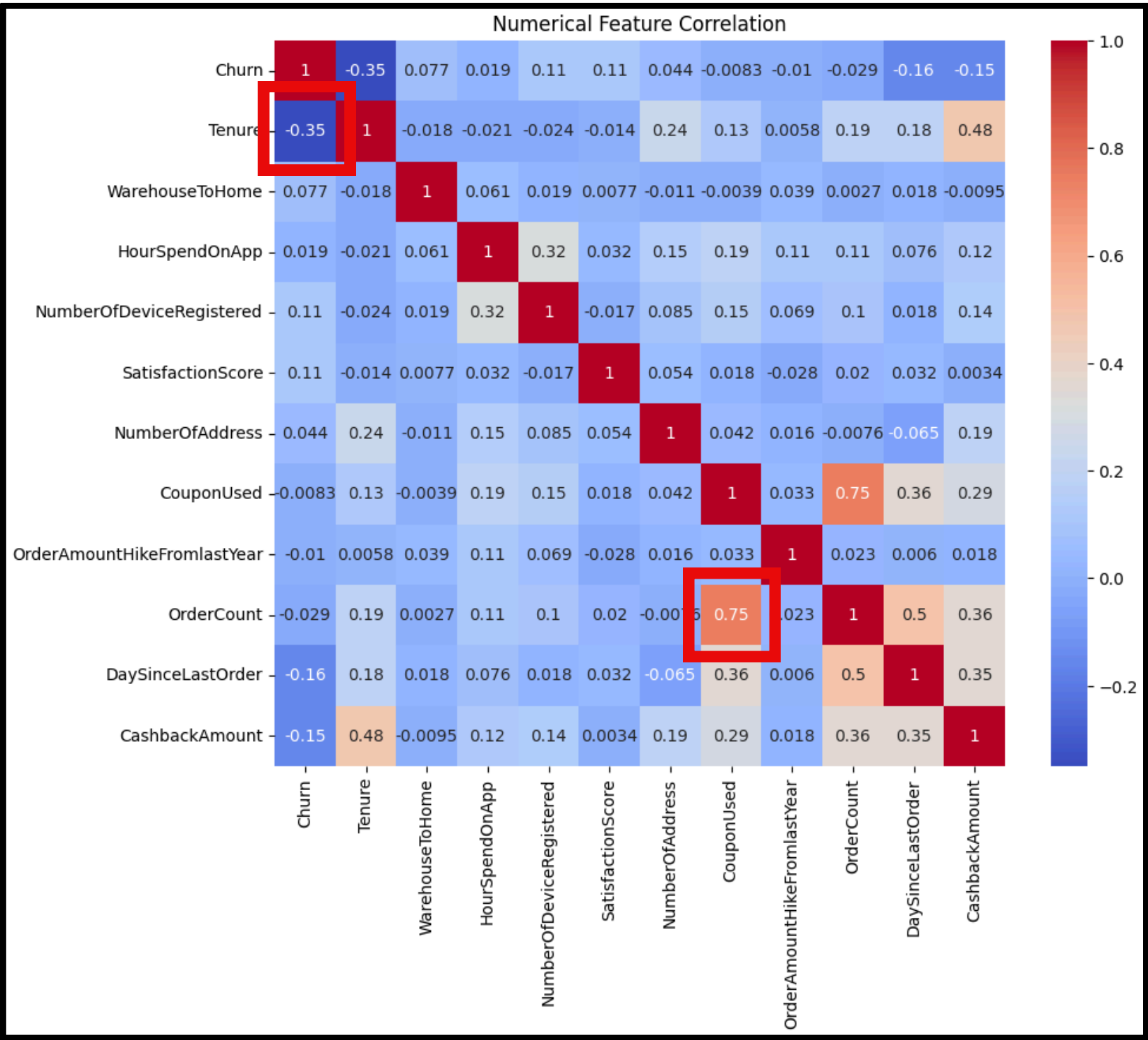
用戶婚姻狀態



分析

大多數用戶為已婚人士，而單身的用戶流失率明顯高於其他族群

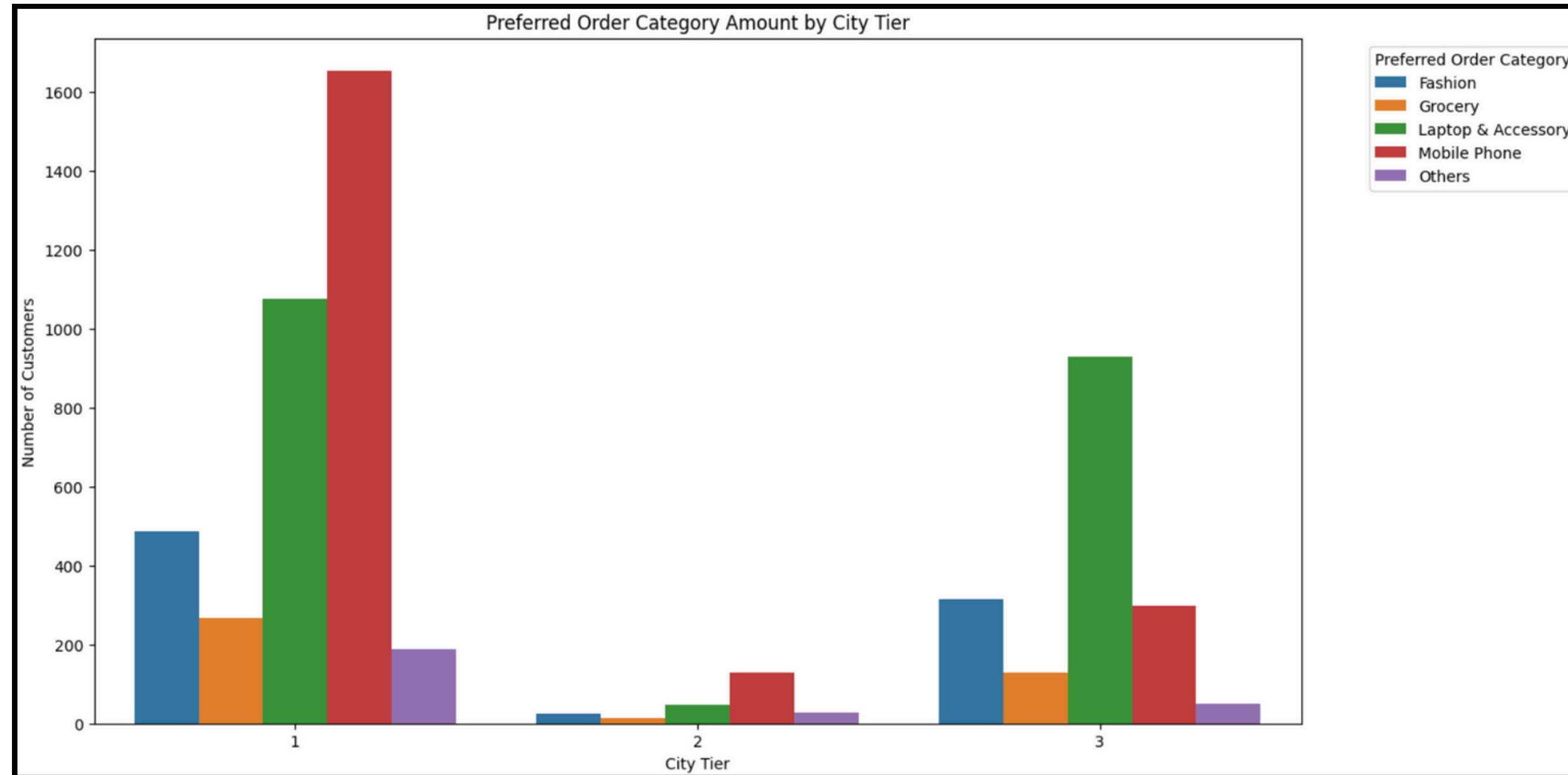
數值型欄位關聯



分析

- 1. 用戶在平台上的活躍時長(Tenure)與流失率有較高關聯性
- 2. 上個月訂單數(OrderCount)與優惠卷使用數(CouponUsed)有高度正相關，為避免多重共線性，後面特徵工程部分會將與流失率關聯較低欄位(優惠卷使用數)刪除


各城市商品類別偏好



分析

1. City Tier 1 & 2最熱門的商品類別為手機(Mobile Phone)
2. City Tier 3 最熱門的商品類別為筆電及週邊(Laptop & Accessory)

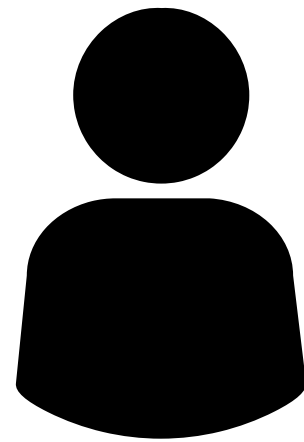
各模型分數

模型	精確率 precision	召回率 recall	f1-score
Logistic Regression	0.48	0.84	0.61
Random Forest	0.95	0.83	0.88
XGBoost	0.97	0.92	0.94
 MLP	0.96	0.95	0.95

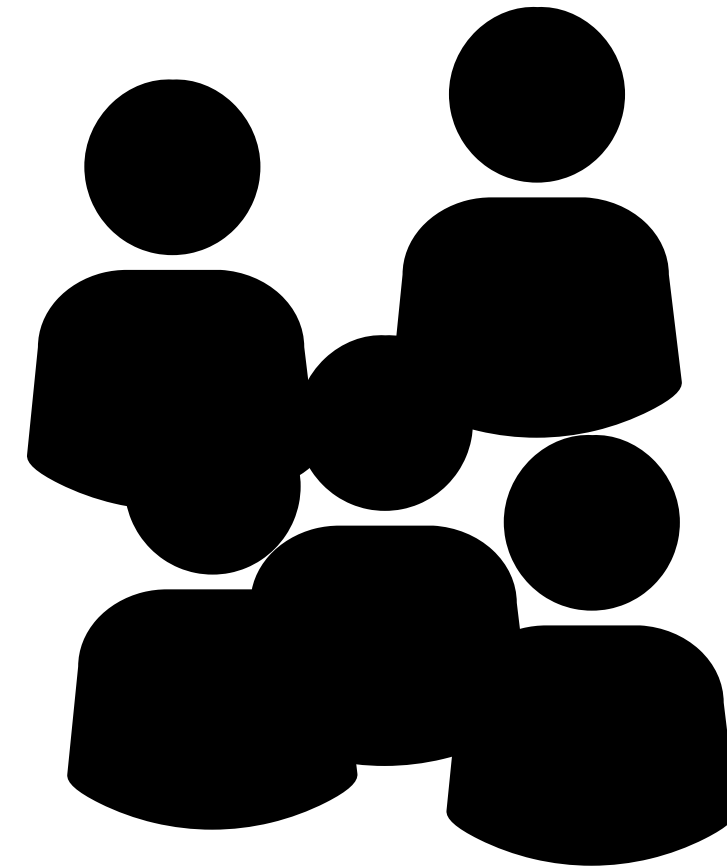
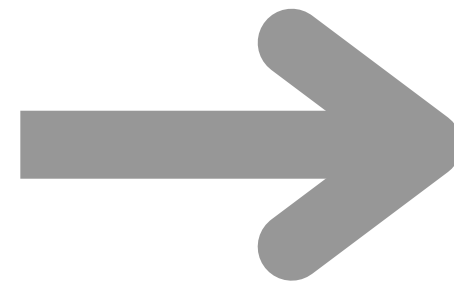
分析

因在流失率分析中，我們較在意recall（漏網率），我們不希望漏掉真正會流失的客戶，所以選擇recall最高的模型
MLP當作最後模型

主要挑戰



因為訓練資料集中流失顧客佔少數，導致模型會偏向永遠預測主流類別（留存客戶），在預測流失顧客中較容易失準



使用SMOTETomek方法，增加樣本數。先使用SMOTE方法，在少數類別（例如流失客戶）合成新的樣本，增加它們的數量。再使用Tomek方法，找到邊界模糊的資料點（互為最近鄰但不同類別的點），刪除雜訊資料，清理決策邊界。因流失顧客樣本數提高，模型更容易預測流失顧客

未來應用



客製化行銷

針對即將流失客戶，設計客製化行銷活動 ex.專屬折扣券、客製化推薦商品。



登入提醒

可搭配限時折扣、登入即送獎勵等激勵手段激活即將流失顧客