
分析報告

背景:

Cyclistic 是位於芝加哥的共享單車公司，提供各種不同類型的單車，包括普通自行車、三輪車，以及適用於殘障人士的自行車。該公司目前提供三種價格方案選擇:單次騎乘票、全天票和年度會員。前兩者適用於非會員(casual)，而最後一個則是 Cyclistic 的會員(member)。

目的:

市場總監認為，大幅增加會員數量能夠最大化公司的利潤，計畫目標不是吸引新客戶，而是吸引非會員成為會員。

關鍵問題:

- 會員以及非會員區別?
- 要如何吸引非會員成為會員?

限制:

由於隱私權保護，數據中沒有個別用戶 id，因此無法追蹤單一用戶並分析他的行為。

資料處理方法:

1. 資料清洗

1. 填補缺失值：如果有缺失值，嘗試在可能的情況下進行填補。在這個專案中，有兩種情況我必須填補缺失值。首先是空的 **GPS** 位置，我通過在數據集中查找相同的站點名稱和 **ID** 來填補它們。其次是空的站點名稱和 **ID**，我通過在數據集中查找相同的 **GPS** 位置來填補它們。
2. 數據驗證：驗證數據集的每一列，檢查是否處於正確的格式，如果不正確，則進行更正。在這個專案中，只有兩種情況需要修復，一種是當兩個站點具有不同的 **ID** 但位於相同的 **GPS** 位置時，另一種是當兩個站點具有相同的 **ID** 但具有不同的 **GPS** 位置時。

2. 資料探勘

嘗試新的特徵來探索和熟悉資料。找尋最有可能導致會員以及非會員差異的原因。而在探索之後，我發現日期和時間是一個重要且值得深入探討的因素。

3. 資料分析

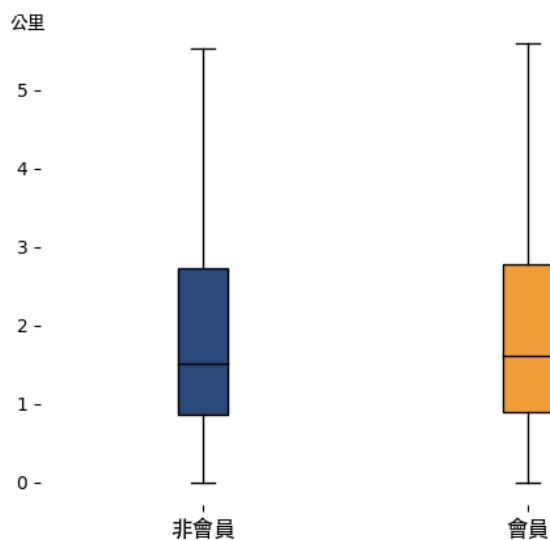
分析資料以及透過資料視覺化尋找會員以及非會員差異

分析:

第一部分

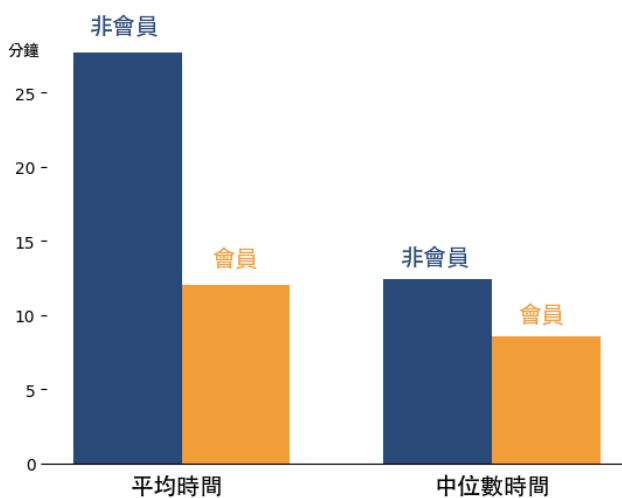
- **距離:** 兩者沒有非常大的差別

會員與非會員的距離分佈箱形圖



- **時長:** 非會員平均下來騎車時間較長

會員與非會員的平均和中位數時間

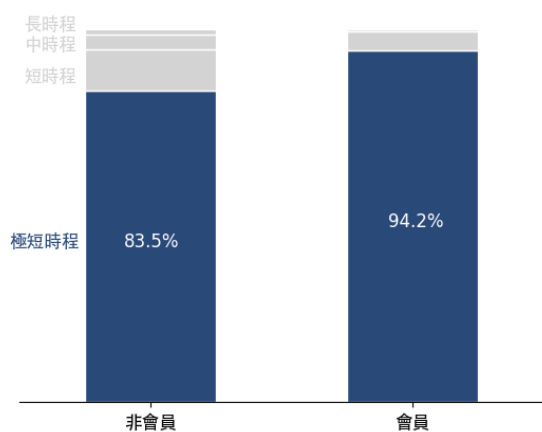


● **時長分組:** 我將時常分為以下四組:

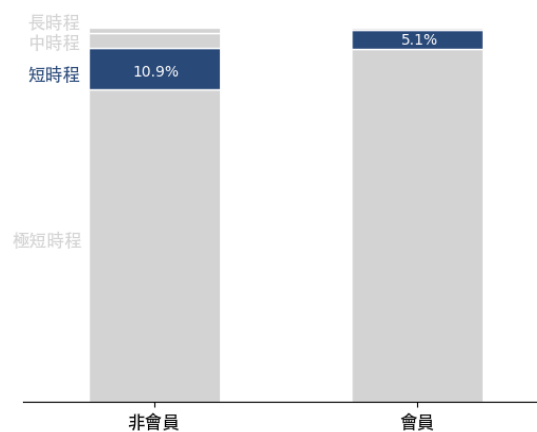
1. 極短時程: 0-30 分鐘
2. 短時程: 30-60 分鐘
3. 中時程: 60-120 分鐘
4. 長時程: 大於 120 分鐘

雖然**極短時程**這類別在兩者都佔了極高的比例，但是**短時程**、**中時程**、**長時程**在非會員卻有較高的比例

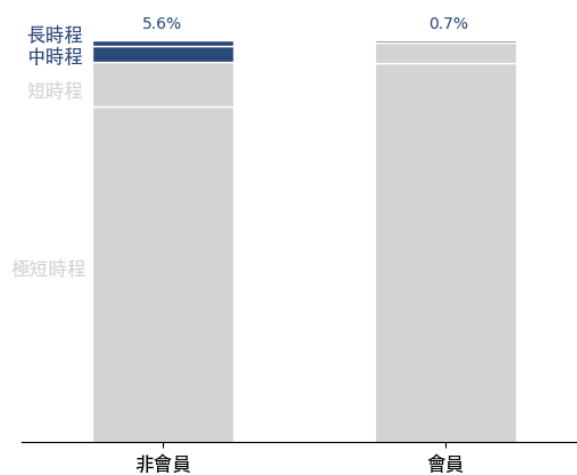
會員與非會員的時程組別比較



會員與非會員的時程組別比較



會員與非會員的時程組別比較



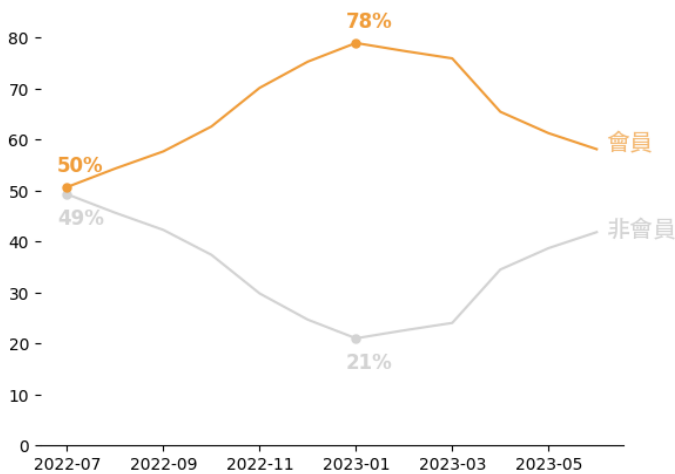
第一部分結論

會員在平均距離相近的情況下花了較少的時間，看起來騎車比較有目的性。而非會員比較像是休閒騎車

第二部分

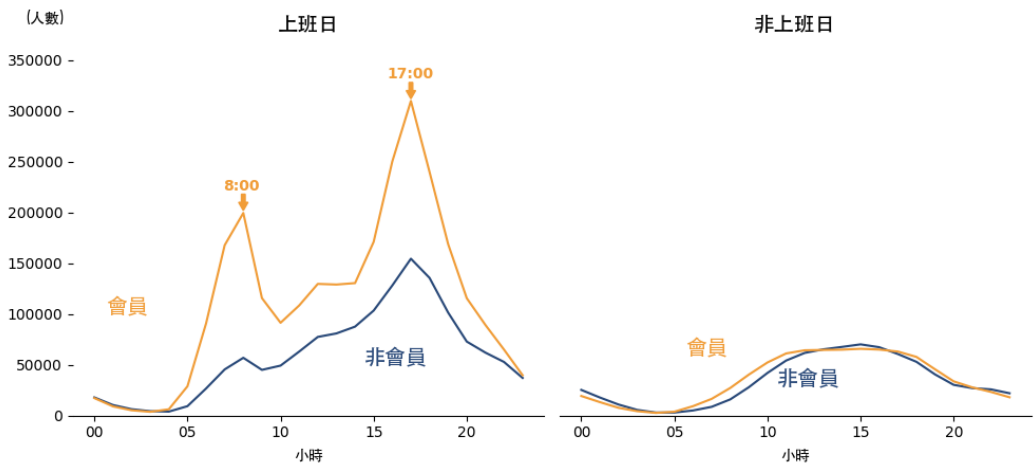
- **會員及非會員在一年中個月佔比:** 夏季時會員以及非會員使用佔比大致一樣，但隨著天氣變冷，非會員佔比急速下滑。一月時達到最低點, 非會員佔比只有 21%

會員與非會員一年中每月所佔比例



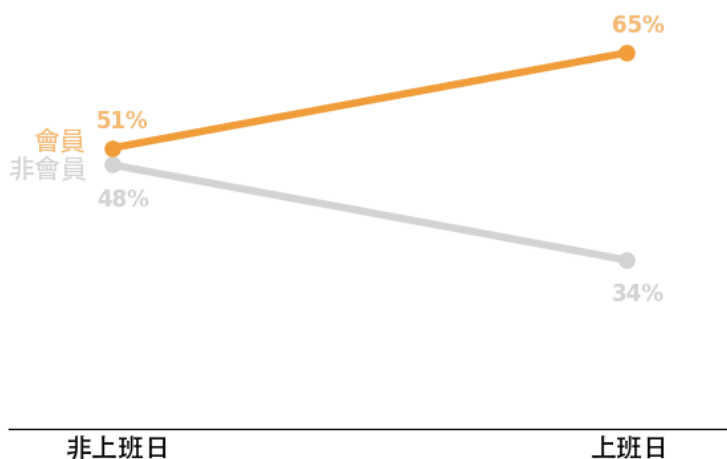
- **上班日使用時段比較:** 在工作天，會員圖上有兩個尖端分別是 8 點和下午 5 點，但到非工作日時兩者幾乎沒有差別

會員與非會員在上班日時段比較



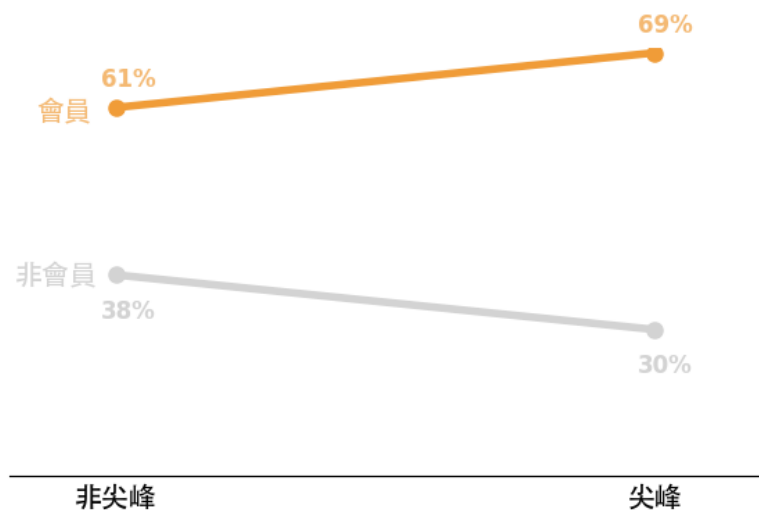
- 上班日: 在上班日, 會員占比從 51% 到 65%

會員與非會員在上班日和非上班日比較



- 尖峰時刻: 在尖峰時刻, 會員占比從 57% 到 68%

會員與非會員尖峰時刻比較



第二部分結論

- 冬天時, 非會員佔比急速下降
- 在工作日和尖峰時段, 會員佔比增加, 可能用來通勤

最終結論

會員:

- 在工作日和尖峰時段，會員更常使用，可能用來通勤
- 會員平均下來騎更快更有目的性
- 幾乎所有會員都騎極短時間(小於 30 分鐘)

非會員:

- 在非工作日，非會員佔比上升
- 非會員騎車較慢較休閒，冬天使用人數急劇下滑
- 有些非會員會騎較長時段(大於 1 小時)

建議:

為非會員制定一個不同的會員計劃：從五月到九月的夏季週末計劃，針對那些休閒騎車的人群。