

1. 專題題目

「代購賣家的定價」或「買家的消費習慣」是否影響買家的下單意願

2. 摘要

最近加入了日本商品的代購社團之後，發現同樣同價的商品，賣家代購的定價會有所不同，所以對某特定商品（以下統稱 A 商品）進行爬蟲，抓取相關資料統計出供需曲線。原本預期會如一般市場供需，但結果與預期不符。

近一步分析其他因素可能會影響買家的需求。因為代購不像一般的買賣，下單後需要等待的時間較長，遇到詐騙的風險也高，有些買家會傾向找之前交易過的賣家代購較為保險，儘管他開的價格較其他賣家來得高。所以除了定價的部分外，我想另外再針對該賣家曾經交易的數量與賣家接到的 A 商品的下單量進行相關性分析。結果可見，兩者相關係數非常高，可見其高度相關。

3. Data:

A. 供需曲線

對從 Facebook 抓下來的原始資料（檔名 data_original.txt，包含發文者名稱、內文標價、留言數＝訂購數），先進一步地進行篩選（檔名 data.txt，已去除不是 A 商品的文章），且因為 fb 有隱藏長

篇貼文的機制 (...查看更多的按鍵)，這點我目前無法寫出更好的 code 來解決這個問題，所以部分被隱藏的資料我也去找原始貼文並補齊。

接著用 excel 對金額與留言數進行初步的整理，包括排序等，並匯出成 csv 檔（檔名 data.csv）以便接下來畫圖使用（檔名 draw.py）。

B. 賣家成交件數是否影響買家購買意願

透過 data.txt，抓取發文者姓名，以「亂數排序取前十」作為抽選的方式，從 Facebook 抓下原始資料（檔名 data_original2-1.txt & data_original2-2.txt，包含發文者名稱、內文、留言數=訂購數），且進行篩選（檔名 data2-1.txt & data2-2.txt，已去除不是代購、揪團的文章）。

接著用 excel 對發文者與留言數進行初步的整理，並會出成 csv 檔（檔名 data2.csv）以便接下來進行相關係數假設檢定。

4. 研究方法：

A. 供需曲線

Step1、以 selenium, BeautifulSoup 對 A 商品進行爬蟲，分別抓取發文者、內文及留言數。（檔名 t.py，這部分可能要請老師加入該社團才有辦法重現程式碼）

Step2、處理原始資料。

Step3、以 matplotlib 來繪製供需圖表（折線圖）。(檔名 draw.py)

B. 賣家成交件數是否影響買家購買意願

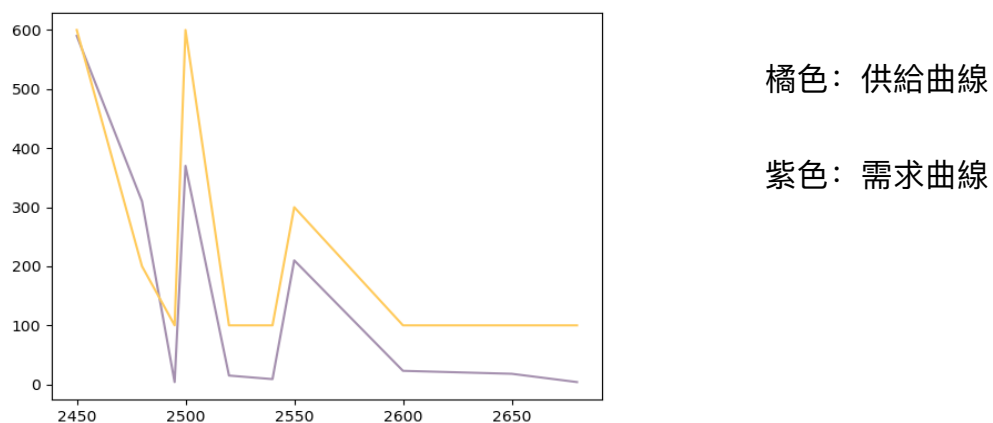
Step1、先以 random 函數選出 10 位不同的賣家，並以 selenium, BeautifulSoup 對其發過的貼文進行爬蟲（僅抓取 10 頁貼文內容），分別抓取內文及留言數。(檔名 tt.py)

Step2、處理資料。

Step3、用 R 語言完成相關係數假設檢定。(檔名 main.R)

5. 結論

A. 供需曲線：



需求曲線還算正常，因為買家會比較賣家所開的價格，選擇與誰進行交易，所以價格越高，需求越低。但其中有些價高者的需求量比價低者還多，推測為買家傾向於向熟悉的賣家訂

購。

而供給曲線就不像一般所認為的價格越高，供給越多。猜想因為代購是比較特殊的產業，有需求才有供給，也就是說賣家通常不會再沒有需求的情況下，進貨商品。且像這樣的粉絲型交易社團，代購通常只是想要降低國際運費的負擔，因個人負擔空運運費會較為昂貴，所以售價開越低，越多人跟團，運費負擔越低。再加上交易平台有規定售價不可高於原商品定價 1.5 倍的規定，故幾乎沒人開高價代購。

B. 賣家成交件數是否影響買家購買意願

假設 $H_0: \rho = 0$ ，及母體相關係數為 0，預期之後會得出一個顯著的相關來否定 $H_0: \rho = 0$ 。

Pearson's product-moment correlation

```
data: data$average and data$A
t = 8.7947, df = 8, p-value = 2.195e-05
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8046599 0.9888799
sample estimates:
      cor
0.9519799
```

由結果可見，兩者間的相關係數為 $r = 0.95$ ，為高度的正相關。且透過檢定的結果，可以拒絕 $H_0: \rho = 0$ ，也就是說，賣家之前的交易量與賣家接到的 A 商品訂購量呈現高度正相關。推斷買家傾向找交易過的賣家交易的結論可能正確。