

專題研究-訂房網站定價策略分析

目錄

1. 參賽感想.....	3
2. 創意說明書.....	4
3. 目前最新進度.....	25
4. 參賽證明.....	26

參賽感想

我在國中開始接觸程式，發現自己對於將資訊應用與現實生活結合的議題有興趣，因為在網路上看過訂房網站相關介紹，想到常常訂飯店時可能因為不知何時訂房而成為冤大頭，因此決定製作程式來分析飯店房價來造福消費者，所以決定以「OTA訂房網站定價策略分析」作為旺宏科學獎參賽題目。

在準備參賽資料的過程中，洪允東老師指導我要根據主辦單位所要求的格式來撰寫創意說明書，並用流程圖將主要觀念解釋清楚。在來回修改內容的過程中，我逐漸學會如何描述問題、從參考資料找出證據支持我的觀點、規劃每個階段性目標與完成的時間。資料蒐集及數據分析這兩部分是這個計畫的主要重點，要爬取什麼資料、多久爬一次，甚至儲存資料的方式，都與後續數據分析息息相關。為了讓評審委員了解計畫的可行性，我先從參考資料找出過往研究影響房價的重要資料項目，規劃資料爬取的頻率與儲存格式，再從網路自學爬蟲程式，藉由比對Hotels.com頁面的資訊來歸納出每個欄位所代表的意義，並把這些重要的概念與流程補充回創意說明書中。在爬取資料的過程中，我學會如何有效的從網路上快速的獲取資訊，並將不同資訊進行交互比對與練習，並進一步運用學習到的知識來解決問題。練習雖花了不少時間，也讓我更熟悉Python語法以及不同爬蟲技巧的優點與限制。

送出創意說明書之後，我繼續爬取後續分析所需要的大量資料，陸續碰到IP被網站封鎖、執行時間過久，程式不穩定，以及每天需要手動抓取資料等一些規劃時沒考慮到的困難問題(詳見目前最新進度)。因此，我除了修改程式，還花更多的時間找同學、學長討論，試著從不同建議中找出正確答案。在這段時間中，執行爬蟲程式蒐集資料已成為我日常生活的一部分，每天為了要保留時間執行程式，我必須利用零碎時間提早完成當天的功課，也慢慢學習到時間管理的重要性。

投入這麼多時間準備，雖然最後連入選資格都沒有，覺得蠻難過的，但我很慶幸自己能在準備比賽的過程中能逐漸確認自己的興趣與努力方向，並透過這次機會學到許多課本上沒教的知識。

第十九屆旺宏科學獎

創意說明書

參賽編號： SA19-362

作品名稱：訂房網站定價策略分析

姓名：鄭恆安

關鍵字：OTA、爬蟲、大數據分析

壹、研究動機

如果你要出遊，如何以有限預算，從茫茫大海中找出理想飯店，是一件非常繁瑣的不可能任務；若你是飯店業者，即使具備吸引顧客的亮點，但缺乏足夠的廣告宣傳，仍可能面對空房率大於入住率的慘況。大家所熟知的Agoda、Expedia、Booking.com等線上旅行社(online travel agency, OTA)，正是因應這個需求所崛起，透過網路提供消費者與飯店業者資訊交流的平台，讓訂房不再是出遊旅行的絆腳石。

對飯店業者而言，上架到OTA能達到良好、免費的曝光管道，但經由OTA銷售出房間所獲取的收入，也會被收取高達房價20%的費用[1]，儘管房價被抽成，也總比什麼都賣不出去要負擔的固定成本來得好。而這些增加的費用，也自然以直接或間接的方式轉嫁到消費者身上。OTA的高額利潤吸引許多新創公司投入此市場，讓飯店業者增加許多選擇的機會。為了增加曝光率，飯店業者可以同時透過多個OTA平台將房間上架，但各OTA平台並沒有經過同步的處理、控管，因此，房間數量有限的小型飯店業者，需要設置通道經理(channel manager)以有效整合各個OTA通路，解決空房、爆房等問題[2]。而房間數量充足的飯店業者，也可能需要有專職的定價經理(rate manager)，以提供更精準的定價策略[2]。為了吸引顧客，部分OTA平台會提供商業情報或數據，像是附近旅宿的平均房價，以利業者做決策。

對消費者來說，各大OTA為了抽取高額的佣金及自身的利益，能否為消費者提供最划算的價格呢？從「你有在網上訂過飯店嗎？有注意到同一個房間在不同的網站上的價格可能不一樣嗎？」這句Trivago深植消費

者的心中的廣告台詞看來，答案顯然是否定的，這些OTA平台為了凸顯各自服務特點的促銷術語，也讓本來對旅行業者及訂房網站上資訊不透明而產生高度疑慮的消費者，產生了極大的信賴殺傷力[3]。為什麼同一個房間在不同的網站上的價格可能不一樣呢？我們都知道，許多商品在快到期前或是乏人問津時，往往有破盤價的出現。反之，如果需求孔急，當庫存不足只剩下最後商品時，物以稀為貴，自然造成價格上揚。研究[4]指出，以國外訂房來說，約60天，也就是提前2個月以前訂房的價格最好，台灣的飯店早鳥優惠則約在45天到30天之前預定即可。但真的是這樣嗎？

如何適當的設定房價？其實是一個飯店業者在OTA興起之前就已經在思考的舊問題，目前也有許多文獻從不同的角度進行研究分析。參考文獻[5] 針對旅館產業住宿服務進行訂價因素與特徵價格之研究，作者應用特徵價格法，探討影響台灣國際觀光旅館房價的因素，文中以旅館的「實體特徵」以及「管理特徵」作為設定特徵價格函數的依據。作者認為可能影響飯店房價的因素有，星級數、是否為風景區、經營定位、規模、交通便利性、淡旺季、地理位置、客群。根據他們的分析結果顯示，房價除了服務成本之外，更隱含景觀與便利性等無形服務特徵的價值。參考文獻[6] 則以臺灣連鎖經營飯店為例，探討電子口碑訊息和房價之間的關係，作者以台灣 50 家連鎖飯店為樣本，利用TripAdvisor 的電子口碑訊息與影響房價因素進行關聯性分析，根據作者的研究結果顯示，飯店規模、地點、舒適、服務等因素是影響飯店房價與口碑的重要因素。參考文獻[7]研究兩岸消費者選擇訂房網站的關鍵因素，作者以問卷調查方式進行分析，研究結果顯示，產品是兩岸消費者選擇訂房網站的最重要的構面，訂房價格、真實性、產品對比和多樣化是台灣消費者選擇訂房網站時的前4項重要關鍵因素。訂房價格、取消成本、會員回饋，和互動性則是影響大陸消費者選擇訂房網站的前 4 個重要關鍵因素。參考文獻[8] 則以網路問卷方式來探討消費者選擇經濟型旅館時，價格與服務品質、硬體設施、地理位置的關聯性。在參考文獻[9]針對旅館業供需管理策略進行探討，本文獻探討的兩個主題之一是「旅館會有淡旺季與離尖峰時段的情形，例如，商務旅館的淡季往往在七月初到九月中，旅館業者如何調整其供需使供需達到平

衡?」作者台中永豐棧麗緻酒店、台北希爾頓大飯店、桃園假日飯店三家國際級飯店為研究對象，以搜集次級資料以及實地個案訪談的方式來進行研究。根據其研究，這三家飯店皆會在淡季時利用折扣來吸引顧客。以永豐棧麗緻酒店為例，折扣在淡旺季時不同，是依據顧客的年住房總合時間分等級，在旺季至多可打 8 折，在淡季(約 6-9 月份)則可打至 6.5 折。台北希爾頓是以商務旅客為主的旅館，因此在星期五、六、日的時候會提供週末假期專案，以較低的價錢和更多的服務來吸引顧客。地處桃園的假日飯店，由於該地區淡旺季之分並不明顯，故其在價格方面的調整幅度不大。以上這些研究，大部分都是根據公、私部門所提供歷史資料的平均值進行分析，雖然從研究中可以看出大概的趨勢，但卻缺乏每天的細部房價，也沒有未來房價變化的細節，因此，相關的結論也無法提供飯店業者所需要的即時定價能力與消費者實際訂房時所需要的資訊。

與現有文獻不同，本研究預定利用OTA訂房網站上所提供的未來房價資料，探討房價主要受那些因素影響?如何訂定策略才能獲得更多的客源?這是本研究希望去探索的第一個問題。既然物價會因供給需求上漲、下跌，飯店業者又會根據各自制定的定價策略來調整房價，哪個網站上的價格會最優惠?原因是什麼?這些價格差異是由於特定OTA平台在競爭下所收取的低廉手續費?還是OTA平台或飯店業者為了吸引特定客群所提出的限時促銷策略?在哪個時間點、哪個OTA平台才是消費者訂房的最佳選擇呢?這是本研究希望去探索的第二個問題。

貳、 研究目的

本研究的目的是利用Python程式，持續自動爬取訂房網站上不同飯店、房型價格，透過OTA訂房網站所提供的長期資料進行大數據分析。針對飯店業者，從每間飯店的定價、調價歷史軌跡，來推敲飯店的定價策略、模式，並推出最適演算法。在初步瞭解每間飯店的定價策略之後，會再進一步研究飯店之間的競合關係，探討每一間飯店以及是否會因應周圍的飯店狀況，甚至是OTA訂房網站的不同促銷策略，動態調整房價。針對訂房者，則探討多久前訂房比較好？如何因應飯店的定價策略或OTA訂房網站的不同促銷策略，訂到最好的價錢？主要研究目標簡述如下：

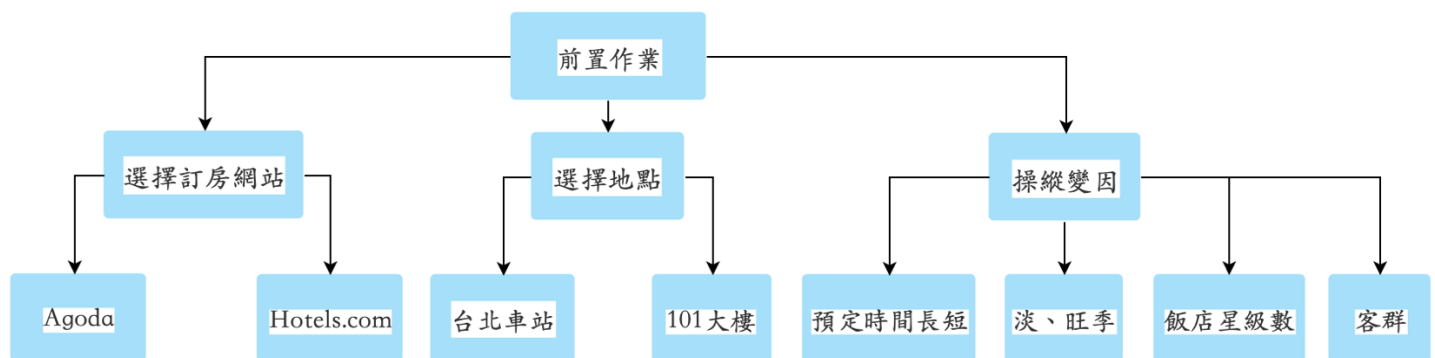
- i. 資料蒐集：長時間從訂房網站上，自動蒐集同一地區，不同訂房日期、飯店星級、房型的價格。
- ii. 數據分析：
 - a. 從飯店業者角度：先從單一飯店開始，研究每一間飯店的定價策略與模式，再從同一地區多間飯店，探討同星級與不同星級飯店間，彼此定價策略與模式的關聯。
 - b. 從訂房者角度：探討如何選擇訂房網站？如何訂到最好的價錢？多久前訂房比較好？OTA訂房網站的促銷活動實際上到底提供多少優惠？

參、 研究方法

為了達到以上目的，本研究預定採用的研究方法，將分成前置作業、以python實作爬蟲程式蒐集資料，以及資料分析、呈現與驗證等三個階段：

一、前置作業

前置作業流程圖如圖一所示，主要分成選擇訂房網站、選擇地點，以及選擇操縱變因三項，相關規劃說明如下：



圖一、前置作業流程圖

i. 選擇訂房網站

	Agoda	booking.com	Hotels.com
價格優勢地區	亞洲	歐洲	無特定
優惠項目多寡	中	中	多
顯示價格是否含稅	否	是	否
刷卡回饋 / 優惠	多	普通	最多

表一、三大訂房網站特色比較[9]

表一是參考文獻[9]所彙整的三大訂房網站特色比較，本研究將根據表一，從國人較常遊玩、出差的亞洲，以及優惠多寡等兩大因素，決定先爬取Hotels.com訂房網站，之後再考慮Agoda，並比較這兩個訂房網站。

ii. 選擇地點

為了要比較不同星級飯店、不同客群對房價所造成的效應，本研究決定選擇台北市交通要衝-台北車站、商業辦公區域-信義區的飯店進行研究。為了要自動爬取資料且避免地址比對判別的複雜度，我採用訂房網站所提供的搜尋功能作為限制條件，以台北車站與101大樓為圓心，自動選取半徑1公里內的飯店進行比較。

iii. 選擇操縱變因

表二是參考文獻[6]所彙整，針對不同地區的飯店進行研究所訂價決定因子，發現現有的參考文獻中並沒有探討訂房預定時間、地理位置、客群、淡旺季、平日假日等因素，也較少提到星級數。因此決定以上述四點進行研究：

- 訂房預定時間(訂房日到入住日的間隔日數)長短
- 星級
- 淡旺季、平日假日
- 地理位置、客群
 - 台北車站(觀光客)
 - 101大樓(商務客)

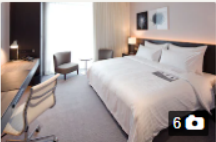
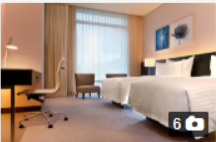
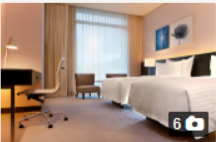
作者(年份)	方法	自變數	應變數	地區
Hung et al. (2010)	一般迴歸與 分量迴歸分 析	規模、營運年數、是否位於風景 區和客房服務員均數	房間定價	臺灣地區國際觀 光酒店(58 間)
Chen and Rothschild. (2010)	特徵價格法	地點、是否有 LED 電視、是否 有開會設施、是否有網路及健身 中心	房間定價 (平日／假 日)	臺北市
Öğüt and Taş. (2012)	迴歸分析	星級數、電子口碑	房間定價和 總銷售量	巴黎和倫敦
Zhang et al. (2011a)	加權迴歸 法、特徵價格 模型	星級數、規模、營運年數、距離	房間定價	北京三星級以上 酒店
Zhang et al. (2011b)	迴歸分析	品質(TripAdvisor網站評價)、地 點	房間定價	紐約
Taso et al. (2015)	結構方程模 式	電子口碑	訂房意圖	臺灣
Ladhari and Michaud. (2015)	實驗設計法	電子口碑(Facebook評價)	訂房意圖	加拿大
盧慶龍等 (2013)	特徵價格法	風景區、都會區、房間數、成立 年數、團客比例、平均每房配置 員工數、國際連鎖品牌	房間定價	臺灣
Mattila. (2001)	實驗設計法	硬體服務品質(食物、設施、氣 氛等)	訂房意圖	北美
Ryu and Han. (2010)	結構方程模 式	軟體(員工服務等)、硬體(食物、 環境等)服務品質	行為意圖	美國
Yang et al.(2011)	Kano Model	服務品質指標、 消費者滿意度評價指標	消費者滿意 度評價指標	臺灣
Browning et al. (2013)	實驗設計法	電子口碑	服務品質	澳洲

表二、飯店訂價決定因子[6]

二、以python實作爬蟲程式蒐集資料

在實作爬蟲程式的過程中，經由不斷嘗試之後，我決定將程式分成「建立房型資料庫」，以及「爬取每天房價」資料兩部分來撰寫：

- i. 建立房型資料庫：透過request模組訪問訂房網站，下載6個月後符合限制條件(台北車站或101大樓半徑1公里內)的所有飯店資料，再用BeautifulSoup進行網頁解析，把每個飯店的每個房型資料建檔。如圖二所示，每間飯店會有多種房型，每個房型會因為服務條件不同而有不同的房價。選擇以6個月後的入住日作為建檔的依

<div>城市景</div>  <div>豪華客房, 1 張特大雙人床 可入住 2 人 (包括最多 1 位兒童) 床型選擇 ■ 1 張特大雙人床 顯示客房資訊</div>	<div>免費無線上網 可於入住時加購早餐 1 位兒童免費入住</div>	<div>Hotels.com™ Rewards ✓ 累積 ✓ 兌換</div>	<div>NT\$7,900 免費取消 2020-07-16 之前</div>
<div>城市景</div>  <div>豪華客房, 1 張特大雙人床 可入住 2 人 (包括最多 1 位兒童) 床型選擇 ■ 1 張特大雙人床 顯示客房資訊</div>	<div>免費無線上網 可於入住時加購早餐 1 位兒童免費入住</div>	<div>Hotels.com™ Rewards ✓ 累積 ✓ 兌換</div>	<div>NT\$7,901 免費取消 2020-07-16 之前</div>
<div>城市景</div>  <div>豪華客房, 2 張單人床 可入住 2 人 (包括最多 1 位兒童) 床型選擇 ■ 2 張單人床 顯示客房資訊</div>	<div>免費無線上網 可於入住時加購早餐 1 位兒童免費入住</div>	<div>Hotels.com™ Rewards ✓ 累積 ✓ 兌換</div>	<div>NT\$7,900 免費取消 2020-07-16 之前</div>
<div>城市景</div>  <div>豪華客房, 2 張單人床 可入住 2 人 (包括最多 1 位兒童) 床型選擇 ■ 2 張單人床 顯示客房資訊</div>	<div>免費無線上網 可於入住時加購早餐 1 位兒童免費入住</div>	<div>Hotels.com™ Rewards ✓ 累積 ✓ 兌換</div>	<div>NT\$7,901 免費取消 2020-07-16 之前</div>

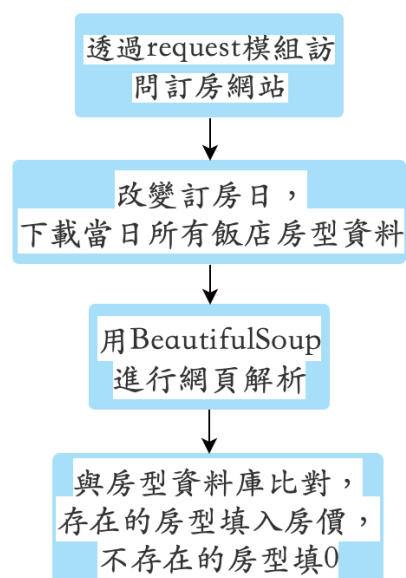
據，主要是避免房間已經被預訂完畢，造成此房型資料不存在，增加後續爬取訂房資料的問題。建立房型資料庫的流程圖詳見圖三。

圖二. 訂房網站上的房型與房價



圖三. 建立房型資料庫的流程圖

- ii. 爬取每天房價：根據建立好的房型資料庫，透過request模組訪問訂房網站，以當日為訂房日，逐日下載入住日從隔日到明年5/6日的每個飯店、不同房型的房價，再用BeautifulSoup進行網頁解析，把所爬取的資料與房型資料庫進行比對，找出不存在的房型，將房價設為定0，代表已沒有空房或是未上架。爬取每天房價的流程圖詳見圖四，程式碼詳見附圖一。

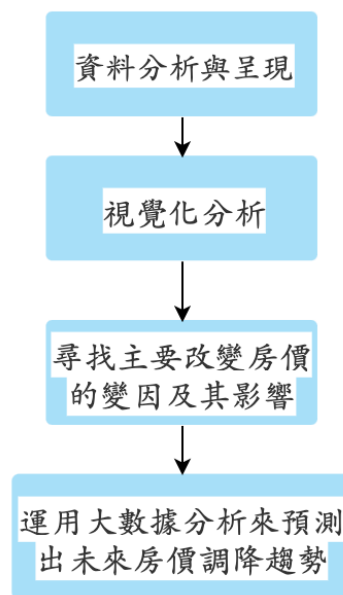


圖四. 爬取每日房價的流程圖

三、資料分析、呈現與驗證

資料分析、呈現與驗證規劃如圖五所示：

- i. 運用excel繪圖，進行視覺化分析
- ii. 根據視覺化分析結果，尋找改變房價的主要因素及其影響
- iii. 運用大數據分析來預測出未來價調降趨勢



圖五. 資料分析、呈現與驗證規劃

肆、初步成果

本研究在選定要分析的訂房網站、飯店所在的地點，以及預定觀察的影響房價因素等前置作業後，就開始以python實作爬蟲程式蒐集資料，進行驗證，並根據爬取的資料進行初步分析。

一、初始環境設定

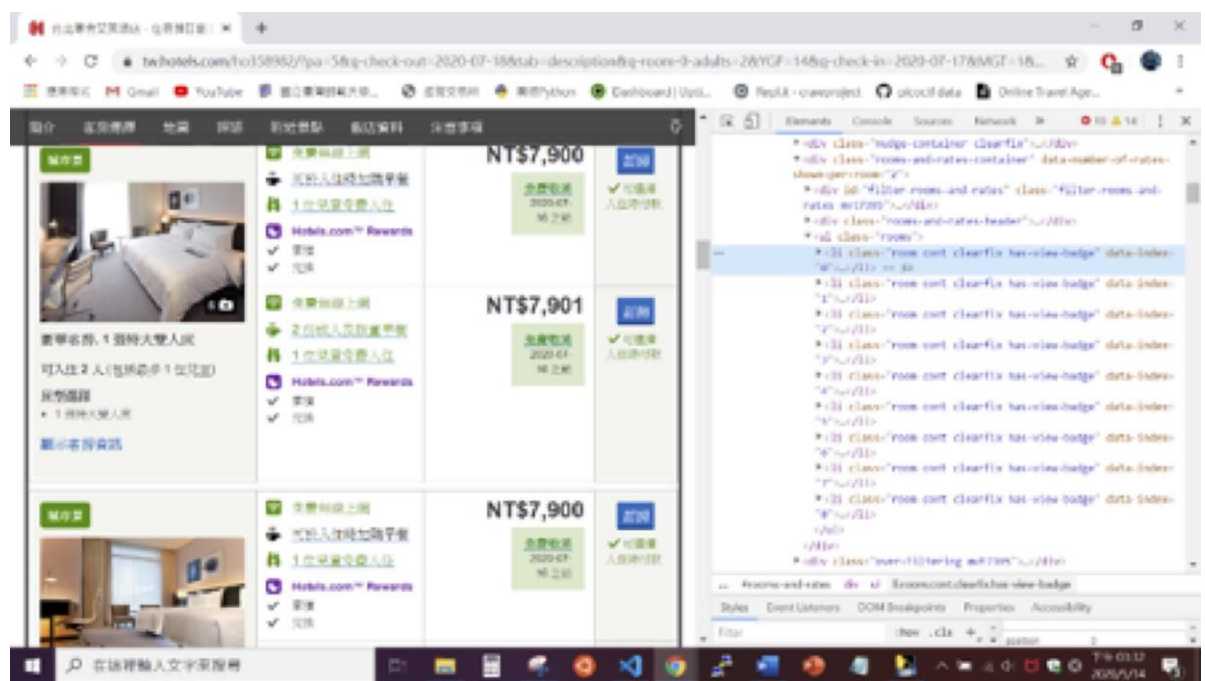
初步先以Hotels.com訂房網站，選擇台北車站4, 4.5, 5星級各一間飯店為例，實作實作爬蟲程式，再將爬取的資料與網站的資料進行比對，以確認資料的正確性。為了探討不同訂房日與入住日的房價的關係，我決定從5/6日起，每天從訂房網站抓取入住日從5/6日到明年5/5日之間每一天的房價。如此，就可以針對任一個入住日，探討不同訂房日的效應。

二、實作爬蟲程式

因為我學過python程式也用過WebDriver實作爬蟲程式，所以在一開始覺得這件工作挑戰並不大，只要上網查程式碼，修改一下即可。但在研究進行的過程中，卻陸續碰到許多之前沒考慮到的問題。例如，一開始我根據網路上的建議，選擇可執行並操作瀏覽器的API介面-Selenium WebDriver，讓程式透過呼叫WebDriver來直接對瀏覽器進行操作完成爬蟲功能，但在初步測試成功，開始大量爬取資料之後，才發現用WebDriver頻繁訪問訂房網頁或爬取太多資料時，會被網站封鎖而爬不到資料。為了解決這個問題，我才找到requests模組來爬取資料。在爬取各飯店房價的過程中，我也發現同一個房間會因為其附帶的服務差異而有不同的價錢，為了降低資料數量以及評估的複雜度，我決定以提供最完備服務選項的房價進行比較。此外，在初步分析爬取訂房網頁所獲得的資料後，發現每天各飯店所呈現的房型資料並不一致，因此常常造成資料排序的錯位，經過仔細比對之後，才發現當飯店沒有空房時，網站會自動捨棄相關資料。為了解決這個問題，我們重新調整實作爬蟲程式的程序，先透過爬取訂房網站一年後入住日期的頁面，先建立每個飯店所有房型的完整資料庫，再於每日爬取訂房網站頁面時加以比對，將沒有空房的房價設為0。最後是資

料存檔方式，一開始我是針對同一飯店建檔，並將各房型皆存入其中，之後改為較容易閱讀的訂房日期對應入住日期所查到的價錢。

在爬取網頁資料後，首先要進行如圖六的網頁解析工作，簡單的概念是透過網頁原始碼，找出網頁中我們有興趣的房型、房價等資訊所相對應的存放位置，以方便後續自動爬取。房型、房價的存放位置分別如圖七、八所示。從圖六，我發現同一個房型依服務項目不同，會訂出不同的價格，但仔細比較之後，發現各家飯店自訂的服務條件不同，不容易進行比較，因此我最後決定相同房型中僅選擇最高房價(選擇服務項目最多)進行建檔。從圖七、八中，我發現每個房型及價錢都在`<li class='room cont clearfix has-view-badge'>`中，說明他們是被放在同一個列表，因此決定用BeautifulSoup 加迴圈來爬取每個房型的房價並建檔。



圖六、網頁解析


```

<div class="rooms">
  <li class="room cont clearfix" data-index="0">
    ::before
    <div class="room-info" aria-labelledby="rr-header-room-type">
      <div class="room-images-and-info">
        <div class="room-image-wrap multiple-images">...</div>
        <h3>
          <span class="room-name">豪華客房，1 張特大雙人床</span>
        </h3>
        <div class="room-and-hotel-info">...</div>
      </div>
      <p>...</p>
    </div>
    <div class="rateplans" role="grid">...</div>
    <div class="additional-room-info">...</div>
    ::after
  </li>
  <li class="room cont clearfix" data-index="1">
    ::before
    <div class="room-info" aria-labelledby="rr-header-room-type">
      <div class="room-images-and-info">
        <div class="room-image-wrap multiple-images">...</div>
        <h3>
          <span class="room-name">高級客房，1 張特大雙人床</span>
        </h3>
        <div class="room-and-hotel-info">...</div>
      </div>
      <p>...</p>
    </div>
    <div class="rateplans" role="grid">...</div>
    <div class="additional-room-info">...</div>
    ::after
  </li>
  ...
</div>

```

圖七、房型位置

```

<li class="room cont clearfix has-view-badge" data-index="0"> == $0
  ::before
  <div class="room-info" aria-labelledby="rr-header-room-type">
    <div class="room-images-and-info">...</div>
    <p>...</p>
  </div>
  <div class="rateplans" role="grid">
    <div class="rateplan" role="row" data-rateplan-id="244091511" data-filter-options="FREE_CANCELLATION|ROOM_HAS_VIEW|DOUBLE_BED_OR_LARGER|EEN_BED_OR_LARGER|FREE_WIFI">
      <div class="info">...</div>
      <div class="pricing">
        <div class="offers-and-prices">
          <div class="prices" aria-labelledby="rooms-and-rates-header-prices" role="gridcell">
            <div class="price">
              <strong class="current-price">NT$7,900</strong>
            </div>
          </div>
        </div>
      </div>
    </div>
  </div>

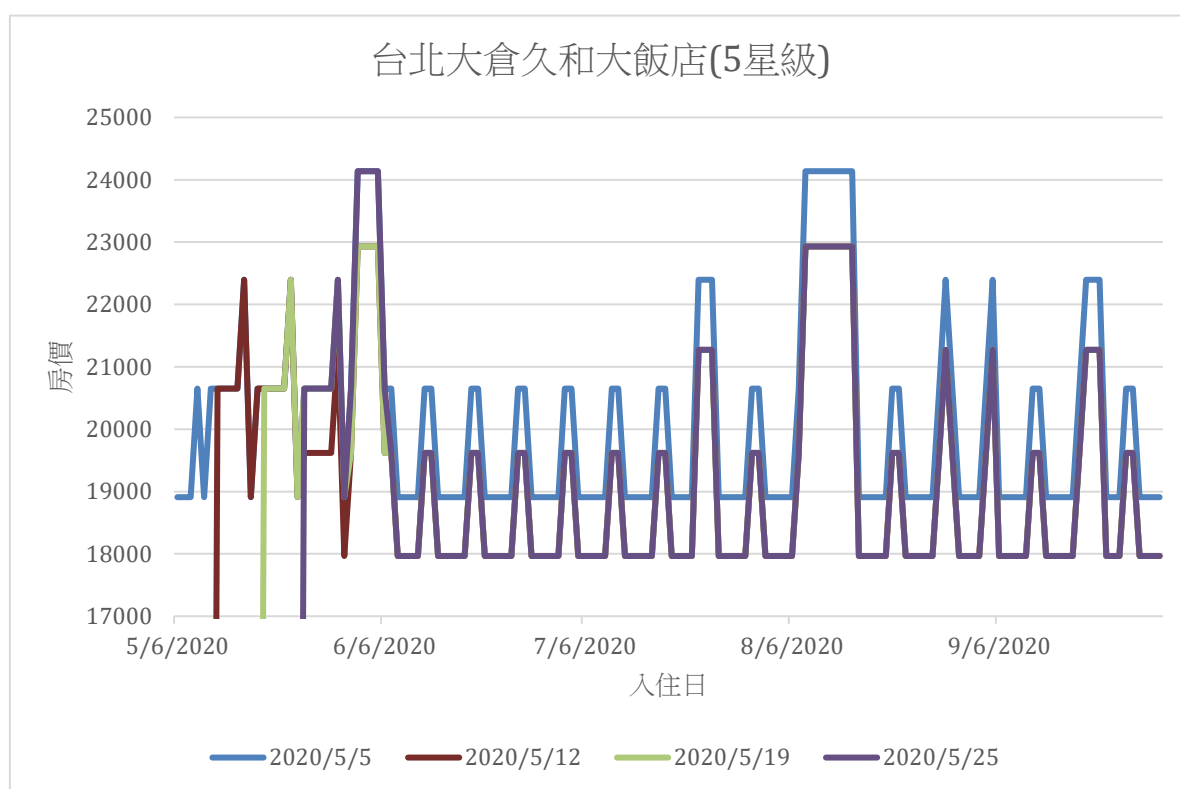
```

圖八、房價位置

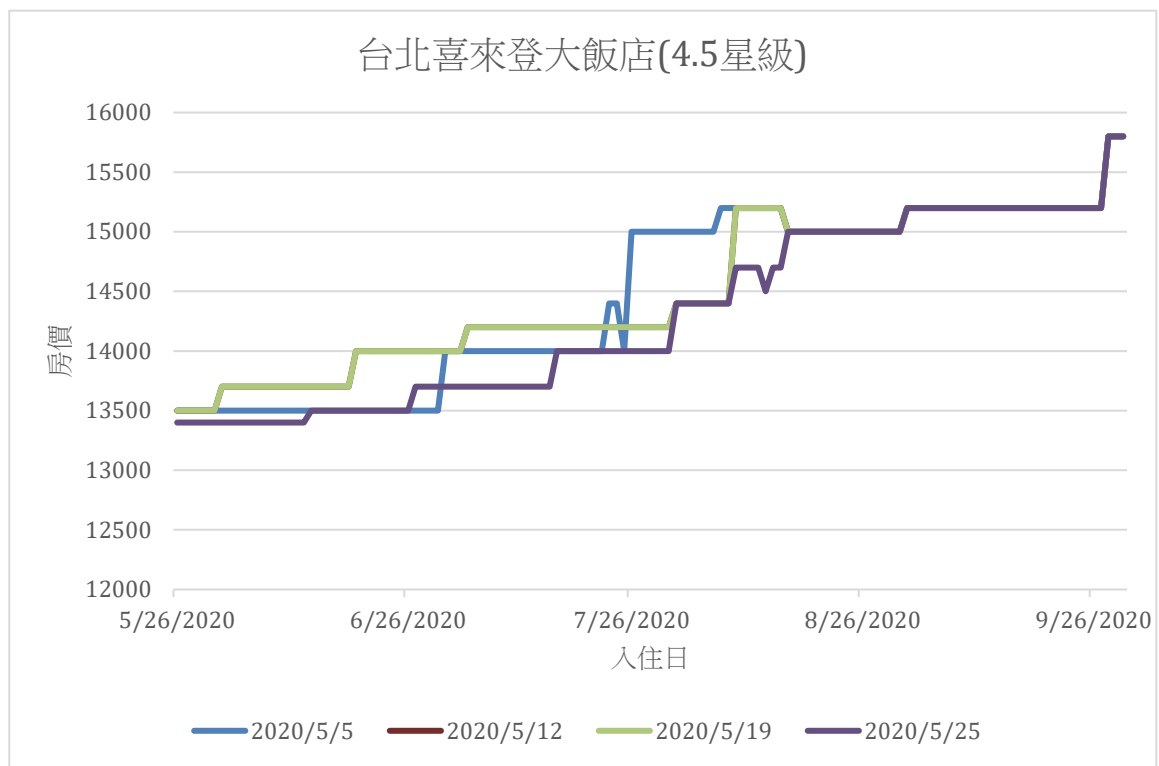
三、資料初步分析結果

經過除錯、修正的過程，我終於可以在5/6日起開始蒐集資料，以下先針對半個多月來的有限資料進行整理、繪圖，並根據圖形進行初步的資料分析，也逐漸透過分析的過程逐漸修正部分資料的錯誤。

圖九到圖十一呈現5, 4.5, 4星級飯店同樣房型的房價與訂房日與入住日關係，其中，X軸代表入住日，觀察範圍是5/26到9/26日；Y軸代表房價，單位是新台幣；每張圖的四條線分別代表5/5、5/12、5/19、5/25這4個不同的訂房日所查詢到的房價。從圖五可發現，在觀察期間內，5星級飯店房價最高價錢與最低價錢相差5000多元，高低價差約為25%。4.5星級飯店房價最高價錢與最低價錢相差2400元，高低價差約為16%。4星級飯店房價的價差為600元，約為14%。可以看出越高價的飯店，價差比例越大。

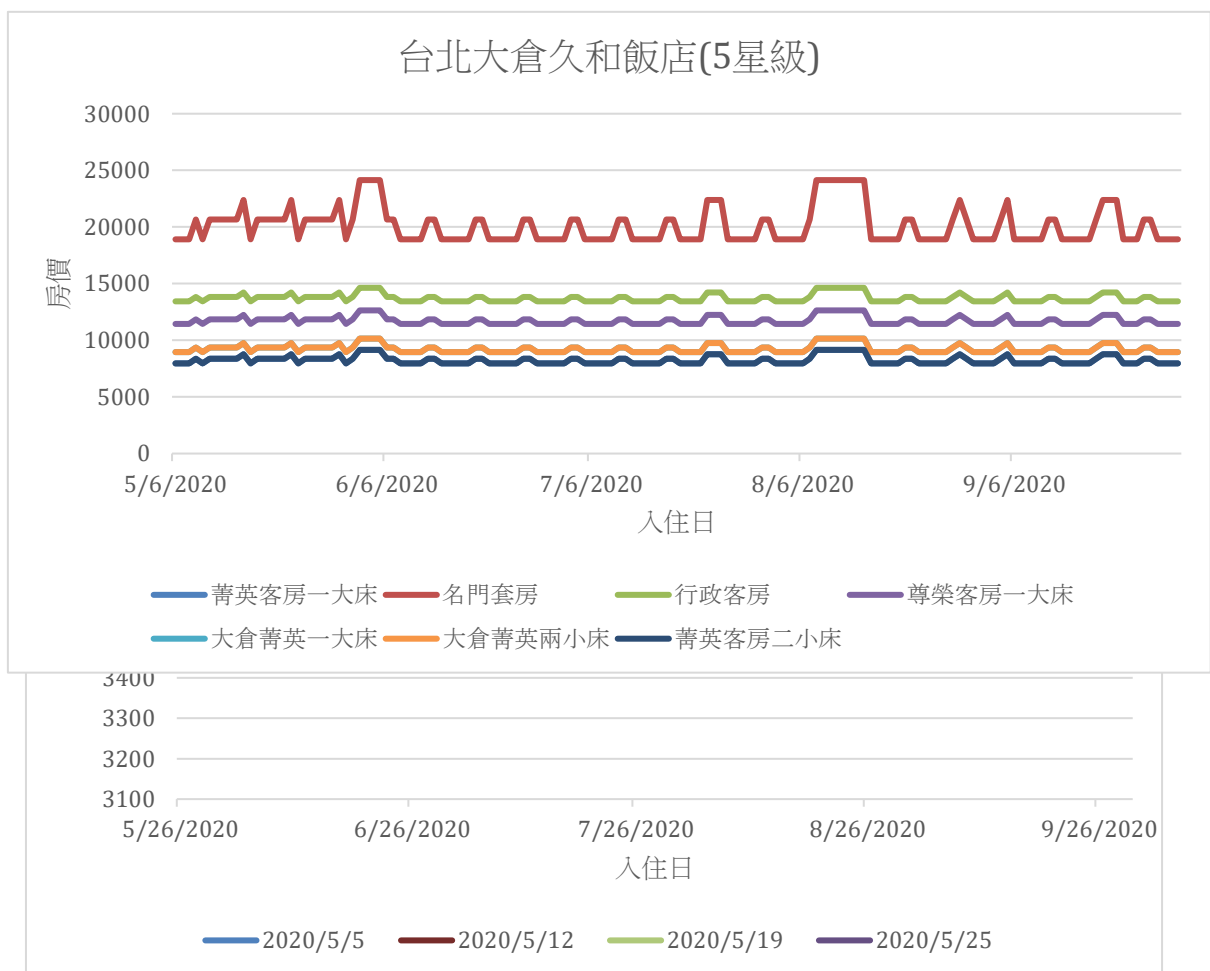


圖九、5星級飯店同一訂房日的房價與入住日的關係

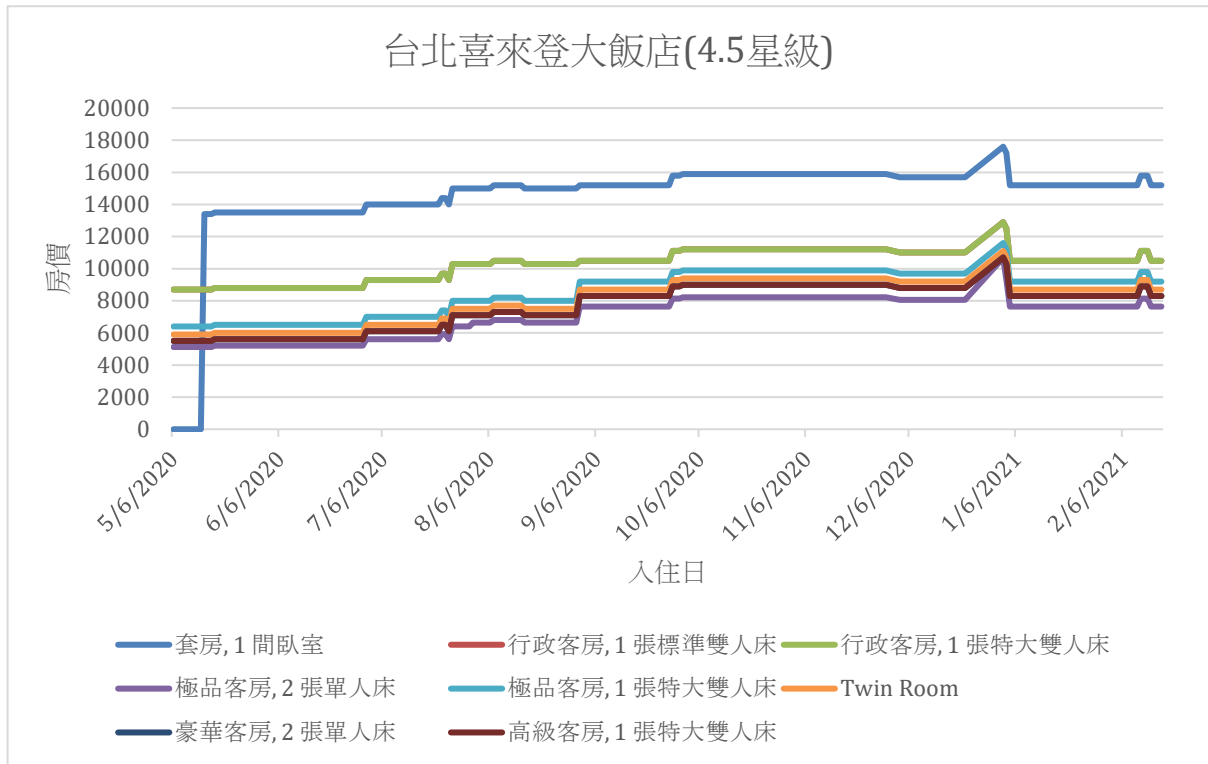


圖十、4.5星級飯店同一訂房日的房價與入住日的關係

圖十一、4星級飯店同一訂房日的房價與入住日的關係

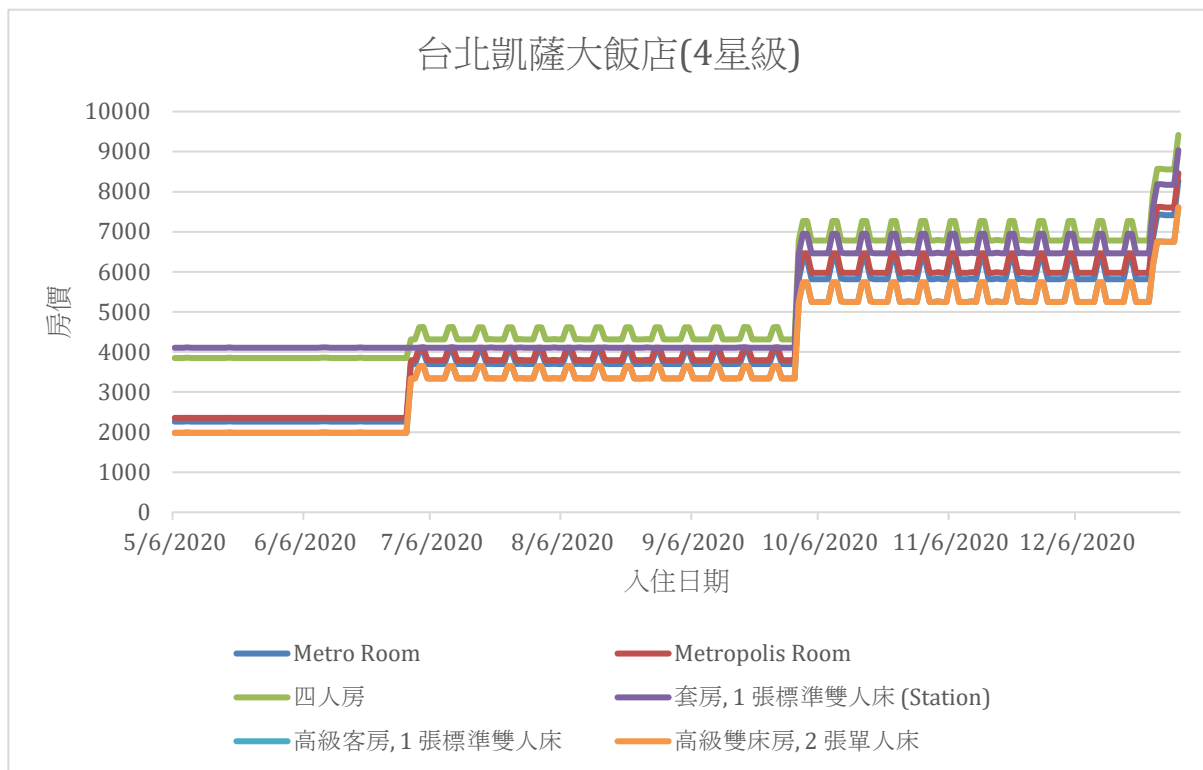


圖十二、5星級飯店不同房型房價與入住日的關係



圖十三、4.5星級飯店不同房型房價與入住日的關係

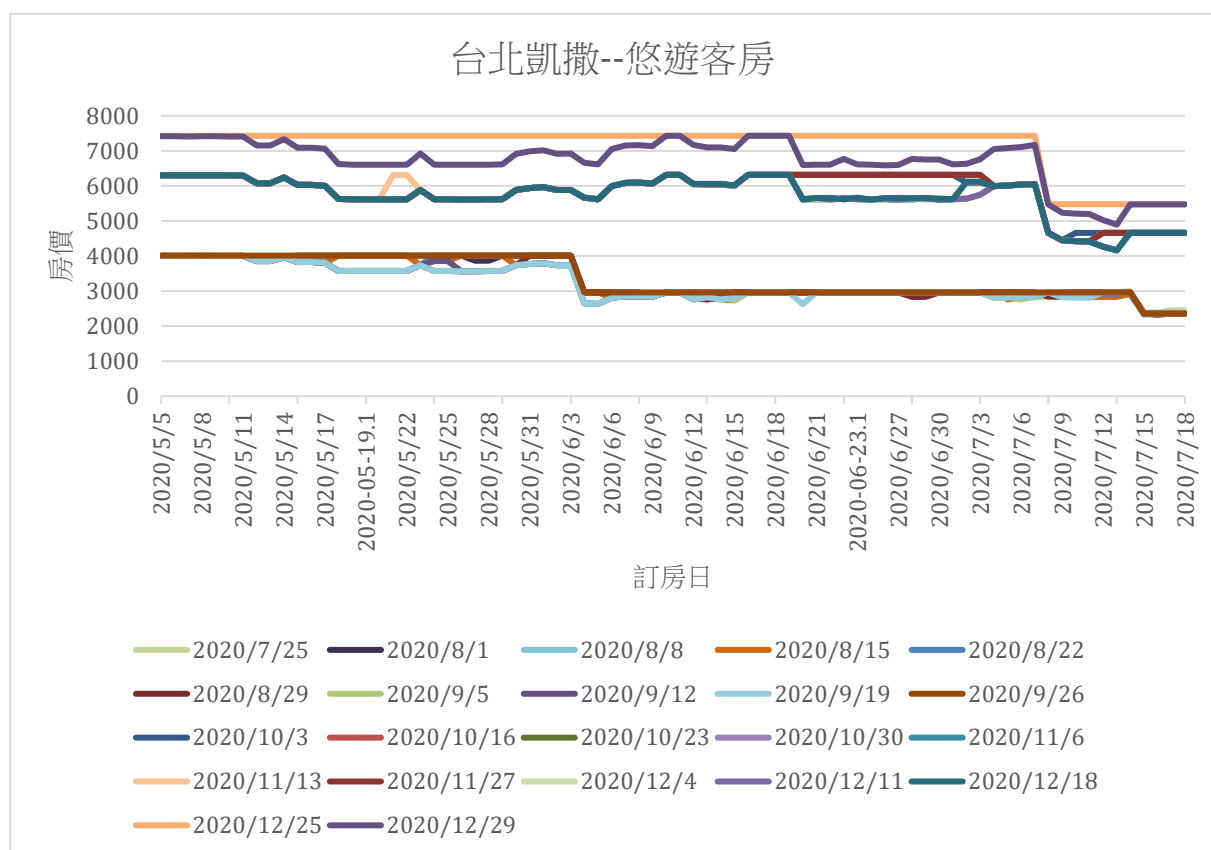
圖十四、4星級飯店不同房型房價與入住日的關係



以圖九的5星級飯店為例，房價的變動採固定的模式，房價基本上持平，並不受到入住日期影響，短期的房價則隨著每週的平、假日上下調整，週末價格較高，平時價格較低。6/2~6日以及8/9~15日這兩個時段應該有特殊的活動，因此價格較高。值得注意的是，5/25日訂房的價格還比5/5低，與過去印象中越早訂越便宜的概念不同，不確定是否因為疫情使飯店決定進一步降價求售，這部分要等疫情穩定後的所取得最新資料才能加以判讀。圖十的4.5星級飯店資料呈現的是另一種完全不同的定價策略，從圖可看出，房價基本上呈現逐步上升的趨勢，可能是業者預期疫情逐漸趨緩，因此逐步上調房價。值得注意的是，5/25日訂房的價錢幾乎都是最低價，5/12與5/19日的訂房價格完全相同，最早的5/15訂房日，卻也在大部分日期都不是最優惠的價格。與圖九類似，入住日於8/9~15日的區間，在5/5, 12, 19日預定時，房價也都是相對拉高。入住日於8/17日之後的房價呈遞增函數，也不受訂房日影響。與前兩圖相比較，圖十一的4星級飯店資料則呈現完全不規律的狀況，最早訂房日(5/5日)所詢得的房價幾乎為定值，與入住日完全無關，且都高於後續三次訂房日的房價，與相關研究所提出的越早訂房可獲得越多優惠的結論並不相同。與圖九相似，5/12日訂房時的房價突然大幅下降，若在8/9日前入住，5/12日訂房時房價均維持為四次訂房中最低。5/19、25訂房的價錢在8/17前入住的房價

又都調漲成最高價，其中，於5/19訂房時，在8/18到9/23日之間會獲得特別優惠，房價為四次訂房中最低。而從圖十二、十三、十四可以發現同一家飯店調價模式大致相同。

根據以上這幾張圖可以看出，各飯店訂價策略差異頗大，但在調整價錢的日期上又似乎可以看出些許關聯性。即使是同一家飯店，在不同訂房日的房價，也看不出比較一致的規律性。而房價調漲、調降的決策過程中，似乎也與過去以歷史資料、問卷或訪談為主的研究所獲得的結論，並不太一致。另外，由於疫情關係造成旅遊人數急速萎縮，所以這段期間所蒐集資料是否足以代表正常時期的調價行為，也是需要再持續觀察的議題。



圖十五、訂房日與房價之關係

從圖十五，可以發現45天前預訂不一定可以獲得最便宜的價格，或許是因為疫情的關係導致飯店的促銷活動，但可從中發現，或許真的有最便宜的時間點，但似乎與預定時間長短不一定相關。

伍、 未來展望

1. 目前只爬取信義區、台北車站各三間旅館，資料量不夠，之後會運程式一次將所有符合條件的旅館都爬取下來。
2. 目前程式是在本地電腦上執行並存檔，因此無法實行每天定時爬取資料。經過觀察發現每經過1小時，網站好像就會更新並小幅調整價錢，因此以目前方式，數據上可能會有些許誤差，之後會將程式放在伺服器或虛擬機上執行、存檔。
3. 架設網站，將資料提供給消費者查詢。
4. 蒐集大量資料後，進行大數據分析，分析主要影響房價的因素，建構出業者調漲價的最適模型。

陸、參考資料

- [1] 如何挑選最適合你的線上訂房平台？(<https://room.fullinn.tw/2019/08/01/ota-choice/?fbclid=IwAR2OTdtnPe1Vla2xpyYDqj3RJrA0GPZUVhIxRCC3C6xxQf7FhtbVXn4f56o>)
- [2] OTA學：時代趨勢下的旅宿產業思維(<https://solomo.xinmedia.com/globaltourismvision/141754>)
- [3] Trivago廣告的成功公式 (<https://www.mrfrank.cc/2019/05/why-the-trivago-video-make-you-deep-impression.html>)
- [4] 如何預訂飯店早鳥房價？(<https://www.skyscanner.com.tw/news/how-to-book-a-cheap-hotel-in-advance-201903>)
- [5] 盧慶龍, 郭曉怡, 陳善珮, “旅館產業住宿服務的訂價因素與特徵價格之研究,” 臺北城市大學學報 第 36 期, 2013 年 05 月
- [6] 侯佩妤, 陳俊智, 包曉天, “電子口碑訊息和訂價：以臺灣連鎖經營酒店為例,” 商略學報, 9卷, 2期, 103-118頁, 2017年
- [7] 李智明, 玄燁, 萬文隆, “兩岸消費者選擇訂房網站關鍵因素研究,” 東吳經濟商學學報, 97期 63-102頁, 2018年12月
- [8] 陳建豪, “消費者對於經濟型旅館選擇之探討,” 碩士論文, 屏東大學休閒事業經營學系碩士班, 2018年7月
- [9] 吳秀玲, 黃煜翔, 黃郁華, “旅館業供需管理策略之探討,” 國立雲林科技大學 企業管理技術系
- [10] 三大訂房網站特色比較(<https://www.moneysmart.tw/articles/agoda-bookings-com-hotels-com-%E8%A8%82%E6%88%BF%E7%B6%B2%E7%AB%99/>)
- [11] 如何避免爬蟲頻繁 requests server 導致被封鎖 IP(<https://python-learningnotebook.blogspot.com/2018/11/requests-server-ip.html?m=1>)

- [12] 查詢-IP-位址是多少(<https://teddybearfp.github.io/2019/09/23/Python-%E6%9F%A5%E8%A9%A2-IP-%E4%BD%8D%E5%9D%80%E6%98%AF%E5%A4%9A%E5%B0%91/>)
- [13] Python 多執行緒 threading 模組平行化程式設計教學(<https://blog.gtwan.org/programming/python-threading-multithreaded-programming-tutorial>)
- [14] python ssl error(<https://xken831.pixnet.net/blog/post/460948442-%5Bpython%5D-%E7%88%AC%E6%96%87%E9%8C%AF%E8%AA%A4-requests.exceptions.sslerror%3A-httpscon>)
- [15] Python SDK for Upstox API(<https://github.com/upstox/upstox-python#listen-to-live-events>)
- [16] Linux 的 scp 指令用法教學與範例：遠端加密複製檔案與目錄(<https://blog.gtwan.org/linux/linux-scp-command-tutorial-examples/>)
- [17] python中的輕量級定時任務排程庫：schedule(<https://www.itread01.com/content/1544596032.html>)
- [18] Linux Crontab (自動化使用)如何操作使用常用，Crontab修改編輯器 記錄(https://dotblogs.com.tw/jason_wang/2016/10/27/crontab)
- [19] 如何使用TorRequests和Python進行匿名請求(<https://www.scrapehero.com/make-anonymous-requests-using-tor-python/>)
- [20] 洋蔥網路學習紀錄(<http://mark1002.github.io/2018/12/01/%E6%B4%8B%E8%94%A5%E7%B6%B2%E8%B7%AF%E5%AD%B8%E7%BF%92%E7%B4%80%E9%8C%84/>)
- [21] Python : requests : 詳解超時和重連(<https://codertw.com/%E7%A8%8B%E5%BC%8F%E8%AA%9E%E8%A8%80/659452/>)
- [22] pandas修改DataFrame列名的方法(<https://codertw.com/%E7%A8%8B%E5%BC%8F%E8%AA%9E%E8%A8%80/358122/>)
- [23] Python中pandas dataframe刪除一行或一列：drop函数(<https://blog.csdn.net/songyunli1111/article/details/79306639>)

- [24] pandas dataframe 删除行和列(https://blog.csdn.net/yyy430/article/details/80810088?utm_medium=distribute.pc_relevant.none-task-blog-BlogCommendFromMachineLearnPai2-1.nonecase&depth_1-utm_source=distribute.pc_relevant.none-task-blog-BlogCommendFromMachineLearnPai2-1.nonecase)
- [25] 如何检查Pandas中是否存在列(<https://cloud.tencent.com/developer/ask/99316>)

```

1  import time
2  import random
3  import datetime
4  from bs4 import BeautifulSoup as bs
5  import pandas as pd
6  import requests
7  import matplotlib.pyplot as plt
8
9  #Home Hotel
10 print('Home Hotel')
11 database = ['普通套房, 1 張特大雙人床', '標準客房, 1 張特大雙人床',
12 '標準客房, 2 張單人床', '高級客房, 1 張特大雙人床, 抗過敏',
13 '高級客房, 2 張單人床, 抗過敏', '行政套房, 1 間臥室'
14 ]
15 #房型資料庫
16 price_total={}#用dict來存取個房型不同入住日的價錢
17 for i in database:
18     price_total[i]=[]
19     today = datetime.datetime.today().date()
20     date = datetime.datetime.strptime('2020-05-24', '%Y-%m-%d')
21     delta = str(today-date.date()).split()#算出訂房日距離05-24幾天
22
23     if(len(delta)==1):
24         t=0
25     else:
26         t = int(delta[0])
27
28 for day in range(1,365):
29     checkin = (datetime.datetime.now()+datetime.timedelta(days=day-t))
30     checkout = (checkin+datetime.timedelta(days=1))
31     checkin = str(checkin.date())#入住日期
32     checkout = str(checkout.date())#退房日期
33     url='https://tw.hotels.com/ho386532/?pa=10&q-check-out='+checkout+
34     '&tab=description&q-room=0-adults=2&YGF=14&q-check-in='+checkin+
35     '&MGT=1&WOE=4&WOD=3&ZSX=0&SYE=3&q-room=0-children=0'
36     html = requests.get(url).text
37     print(checkin)
38     soup = bs(html, 'lxml')
39
40     roomtypes = soup.find_all('li',
41     class_='room cont clearfix has-view-badge')
42     roomtypes += soup.find_all('li', class_='room cont clearfix')
43     #不同飯店網站上class名稱會不同
44
45     print(len(roomtypes))
46     for roomname in roomtypes :
47         style = roomname.find('span', class_='room-name')
48         price=[]
49         if len(style)==0:
50             pass
51         prices = roomname.find_all('ins', class_='current-price')
52         for i in prices:
53             print(i.text)
54             price.append(i.text)
55         prices = roomname.find_all('strong', class_='current-price')
56         for i in prices:
57             print(i.text)
58             price.append(i.text)
59         #將價錢存入各房型的陣列中
60         price_total[style.text].append
61         (int((price[-1].replace('NT$', '').replace(', ', '')))
62         #因為在excel中，逗號代表分隔。
63         #捨棄NT，價錢轉換為數字，才能用程式做比較
64     for k in price_total:
65         if len(price_total[k])<(day):
66             price_total[k].append(0)#處理空房問題

```

附

圖一、程式碼

目前最新進度

一、運用程式一次將所有符合條件的旅館都爬取下來

- i. 實作
 - 1. 先到 Hotels.com 搜尋所有符合條件的旅館
 - 2. 進行滾動頁面，點擊所有旅館，儲存Hotels.com 上每個旅館及房型的資料編號
 - 3. 每次根據旅館的編號、搜尋網站並爬取房價
- ii. 遇到的問題
 - 1. 爬行所花的時間過長導致IP被伺服器封鎖
 - 2. 因為IP被伺服器封鎖，造成資料重複爬取
- iii. 解決辦法
 - 1. 運用Python threading 實施多線程爬蟲解決IP封鎖問題
 - 2. 使用dataframe drop和rename函數解決資料重複爬取問題

二、將程式放在伺服器或虛擬機上執行、存檔

- i. 實作
 - 1. 先在本地電腦爬取房型及旅館編號的資料
 - 2. 上傳爬蟲程式到Container上執行
- ii. 遇到問題
 - 1. 爬取一段時間後就會出現Response<403>，伺服器收到請求但拒絕提供服務
- iii. 解決辦法
 - 1. 利用tor request 每三十分鐘定期更換IP

參賽證明



旺宏科學獎

證書編號：2020-SA19-362

證書

國立師大附中 鄭恆安 同學
具科學研究熱忱並積極參與第十九屆科學獎競賽
及完成繳交作品，特頒此證，以茲證明。

作品名稱：OTA訂房網站定價策略分析

指導老師：國立師大附中 洪允東 老師

財團法人旺宏教育基金會

董事長

吳敏求



中華民國一〇八年七月十五日