

國立臺灣師範大學附屬高級中學第47屆科學展覽會

作品說明書封面

科 別：資訊

組 別：高中組

作品名稱：訂房網站定價策略分析

關鍵詞：OTA、爬蟲、LSTM

編 號：

摘要

住宿是安排旅行行程重要的關鍵之一，大部分的人都希望能用划算的價錢訂到心目中理想的飯店，但要達到這個目標，就需要了解每間飯店的訂價策略。本研究根據過去的參考文獻，並利用線上訂房網站上取得的大量房價資料，透過與數據分析找出影響房價的關鍵因素，再利用這些關鍵因素與蒐集到的大數據去訓練LSTM模型，並用訓練好的LSTM模型來預測未來房價。

找出在哪個時間點才是消費者訂房的最佳選擇。

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，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[1]。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。以網路旅遊巨頭Expedia為例，其自行開發一套資源開發營收管理工具「Rev+」，這套系統藉由機器學習技術，依照各市場的條件，分析與競爭對手過去90天內的房價、住房率等數據，接著與整個Expedia旗下品牌的預訂數據整合，得出未來的市場供需，每天為業者提出定價建議[2]。

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。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？」作者以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，在旺季時至多可打8折，而在淡季(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？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，OTA又會根據其人工智慧來預估未來客源並調整房價，那在哪個時間點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

本研究的目的是藉由機器學習的技術，應對OTA的調價工具，消費者提供最大利益。為達到此目的，本研究利用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。針對調價的主因，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。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，動態調整房價。最後根據上述結果，歸納出訓練LSTM的特徵(features)，增強模型預測的實用性，進而解決多久前訂房比較好才會最划算的問題。

主要研究目標簡述如下

一、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。

二、數據分析

(一)、分析影響房價的主要因素:

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，以及最後不同地區(客群)多間飯店、探討地區、星級造成定價策略與模式的關聯。

(二)、LSTM預測分析:

根據(一)所得到的主要因素，制定適合的特徵(features)，訓練並觀察模型準確度，並調整參數以達最好預測成果。

參、研究設備與器材

一、硬體：筆記型電腦、可攜式硬碟

二、軟體及工具：Python、Visual Studio Code、Google colab

Python相關套件	功能
requests、webdriver	建立HTTP請求、自動化瀏覽器
BeautifulSoup	用於解析網頁內容
datetime	處理時間和日期
pandas	用於數據操縱和分析
sklearn	提供大量常見的機器學習演算法和許多實用的資料集合
matplotlib	資料視覺化

肆、研究過程與方法

一、定義問題

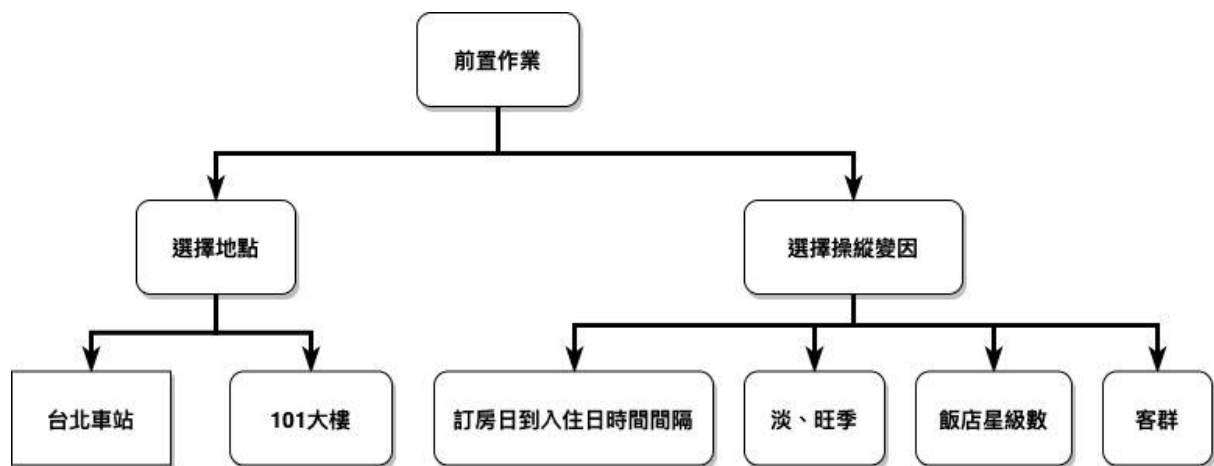
從參考資料[2]，我們發現OTA每天會調整未來尚未售出房間的價錢(以下簡稱房價)。因此，在不同時間搜尋相同條件、相同入住日的房價也會不同。所以本研究的核心問題為在未來的哪個時間點搜尋到相同條件、相同入住日的房價是最划算的。

二、實驗流程

為了解決上述問題，本研究預定採用的實驗流程分成前置作業、蒐集資料及資料分析、呈現、訓練模型、實際預測等五個階段。

三、前置作業

前置作業流程圖如圖一所示，主要分成選擇地點及選擇操縱變因兩項，相關規劃說明如下。



圖一、前置作業流程圖

(一)、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，以台北車站與信義區(101大樓)為圓心，自動選取半徑1公里內的飯店進行比較。

(二)、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日到入住日的時間間隔、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：

1. 訂房預定時間(訂房日到入住日的間隔日數)長短
2. 星級

3. 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

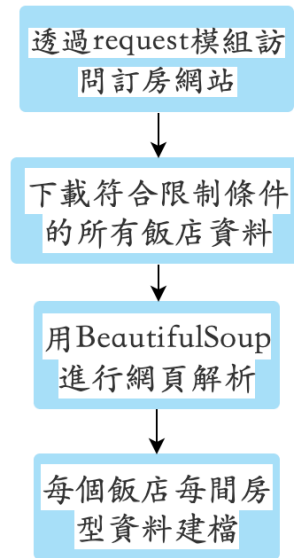
表二、飯店訂價決定因子[6]

四、蒐集資料:

本研究以Hotels.com訂房網站為搜尋的OTA，接著選擇台北車站、101大樓周遭一公里內的3、4、5星級飯店後，再用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分析。而在實作爬蟲程式的過程中，為了處理上的前置工作，之後每天的資料蒐集。因此我將程式分成「建立房型資料庫」，以及「爬取每天房價」資料兩部分來撰寫：

(一)、建立房型資料庫

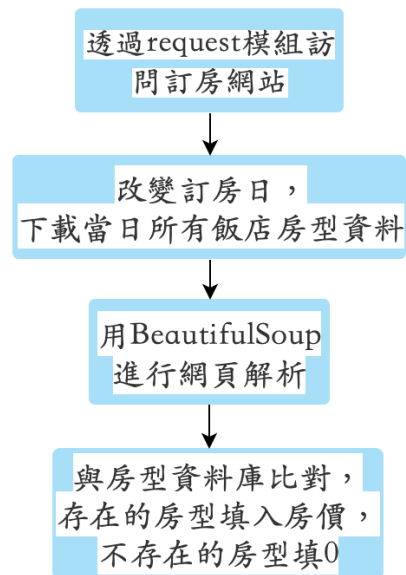
透過request模組訪問訂房網站，下載6個月後符合限制條件(台北車站或101大樓半徑1公里內)的所有飯店資料，再用BeautifulSoup進行網頁解析，把每間飯店的房型資料建檔。如圖二所示，每間飯店會有多種房型，每個房型會因為服務條件不同而有不同的房價。選擇以6個月後的入住日作為建檔的依據，主要是避免房間已經被預訂完畢，造成此房型資料不存在，增加後續爬取訂房資料的問題。建立房型資料庫的流程圖詳見圖二。



圖二.、建立房型資料庫的流程圖

(二)、爬取每天房價

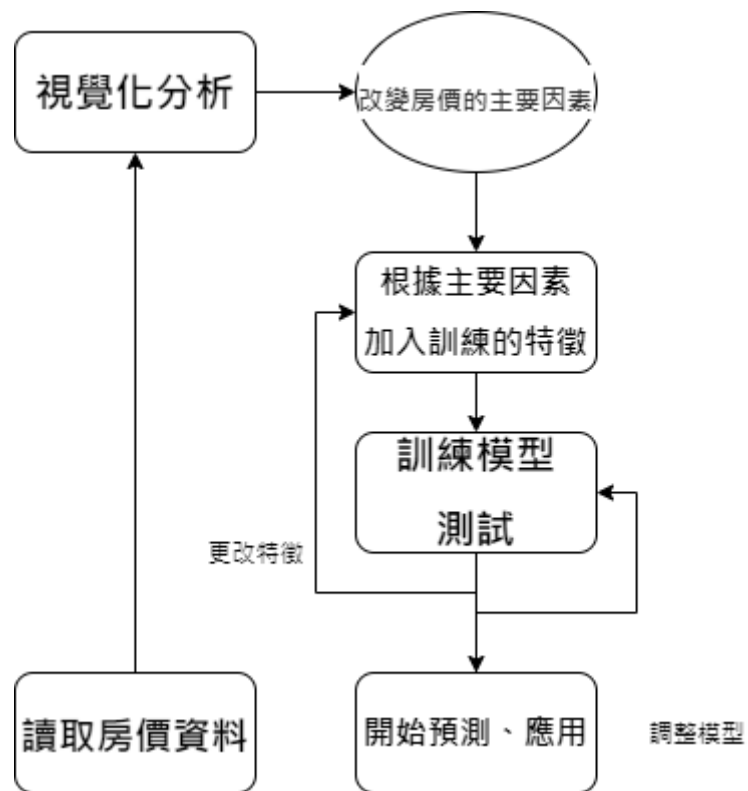
根據建立好的房型資料庫，透過request模組訪問訂房網站，以當日為訂房日，逐日下載入住日從隔日到2021/07/01的每間飯店每個房型的房價，再用BeautifulSoup進行網頁解析，把所爬取的資料與房型資料庫進行比對，找出不存在的房型，將房價設為定0，代表已沒有空房或是未上架。爬取每天房價的流程圖詳見圖四，程式碼詳見附圖一。



圖三、爬取每日房價的流程圖

五、資料分析、呈現

- (一)、運用python matplotlib模組繪圖，進行視覺化分析
- (二)、根據視覺化分析結果，尋找改變房價的主要因素及其影響
- (三)、根據觀察到的因素與房價找相關性，加以判斷其影響程度
- (四)、利用上述觀察到改變房價的主要因素作為特徵(features)，並使用長短期記憶模型(long short-term memory，LSTM)來預測出未來房價走勢



圖四、分析、訓練流程圖

六、 訓練LSTM模型

根據視覺化分析後，我們發現入住日的日期、是星期幾、是什麼節日、搜尋日與入住日的時間間隔以及搜尋日搜尋到入住日的房價，圖五。

七、 實際預測

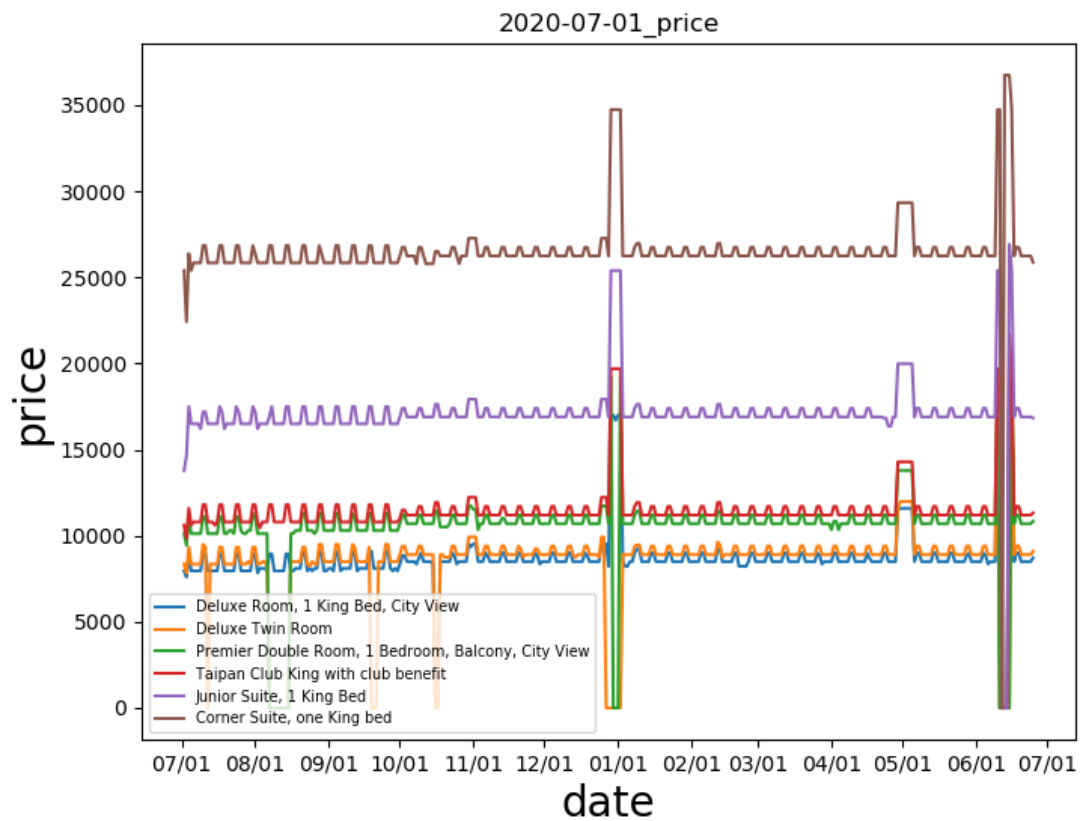
伍、 研究成果

為了找出影響飯店價錢的特徵(features)，制定以下三種操縱變因，分別為不同房型、不同訂房日、不同地區(客群)與不同飯店以探討它們與房價的關聯。

一、觀察同家飯店、不同房型的價格變化

(一)、實驗設計

將2020-07-01所搜尋的不同房型的價格走勢繪在同張圖上



圖五、不同房型的價錢走勢圖

(二)、觀察與發現

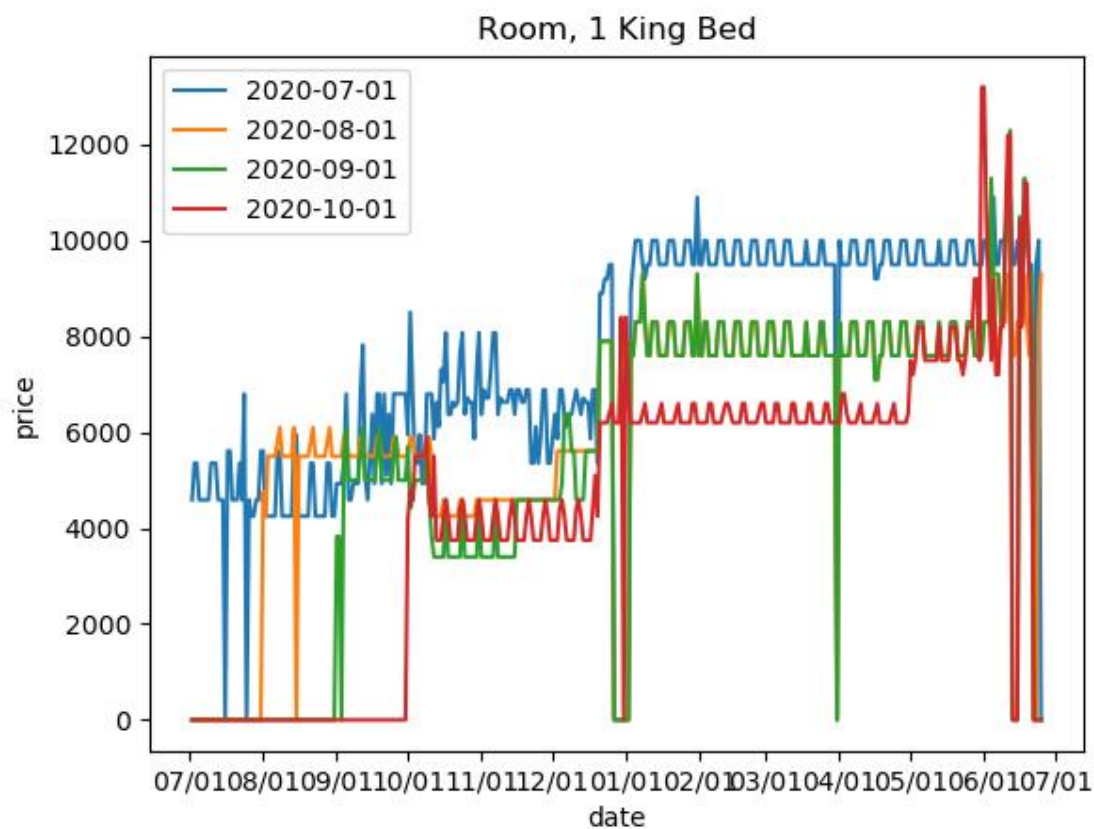
1. 不同房型價格不同
2. 不同房型價格調整有一致的趨勢，經過比對數據後發現，房價會在周五周六兩天調漲，
3. 特殊假日(1/1, 端午連)

由圖五可發現大部分時間皆具有週期性變化，推測為平日、假日之別，少部分時間的大幅調漲應為特殊假期。此外，還可發現不同房型的價錢走勢差異不大，但特別在特殊假期所造成的調價程度並不一致。

二、觀察不同訂房日，相同飯店、相同房型的價格變化

(一)、實驗設計

將訂房日為7/1、8/1、9/1、10/1(搜尋2020/7/1到2021/7/1的每天房價)的四條價格走勢繪在同張圖上



圖六、四條不同訂房日價格走勢圖

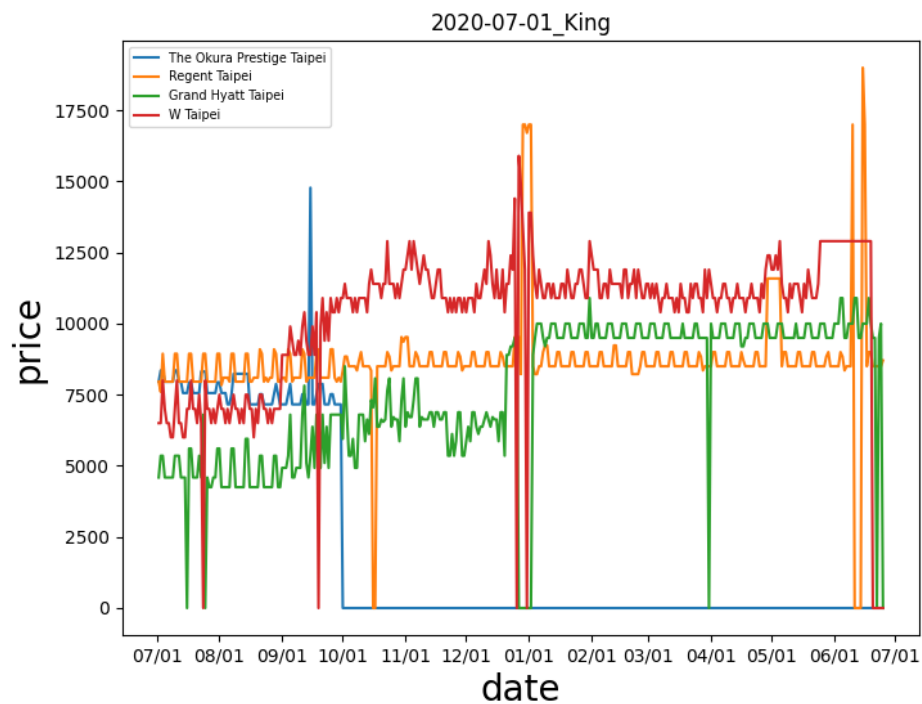
(二)、觀察與發現

由圖六可發現，針對同一房型，隨著訂房日越接近入住日，房價可能會增加，但也可能減少，這跟我們假設的早鳥優惠不同，這也是本專題主要研究的議題之一，更深入研究放在下面的討論。

三、觀察不同地區(客群)、不同飯店、相同星級、相同房型的價格變化

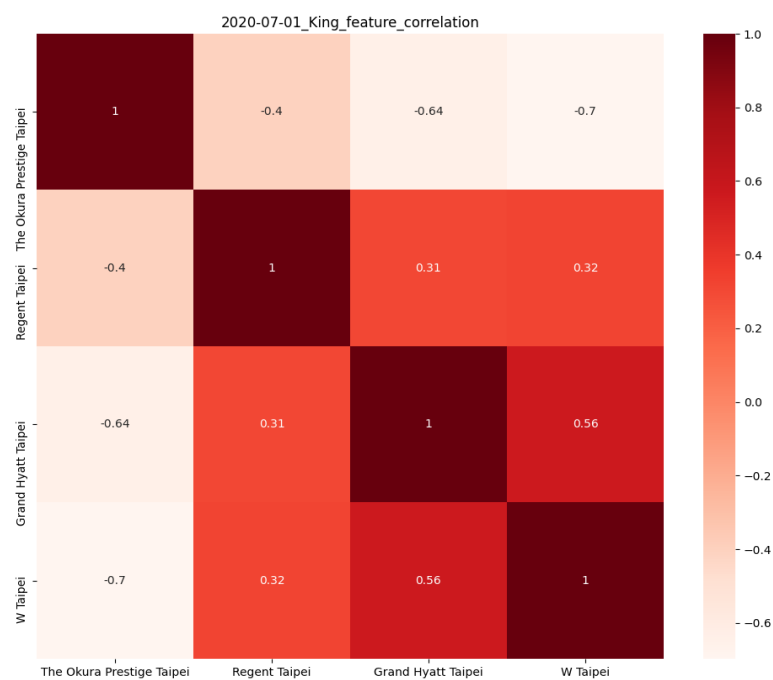
(一)、實驗設計

因為每間飯店的房型名稱及房型等級都大不相同，參考資料[10]歸類出大部分飯店都有的King(單人房)、Twin(單床雙人房)、Double(雙床雙人房)、Suite(套房)等四種房型，因此我先按照各房型的價格由小到大排序，再將每種房型與上述四種房，接著從每間飯店選擇價格最低(最基本)的四種房型，最後以房型為控制變因繪出不同飯店的價錢及之間的相關程度圖。以下以各個星級飯店中最多家飯店上架的房型，分別為五星級的King、四星級的Twin、三星級的Double。



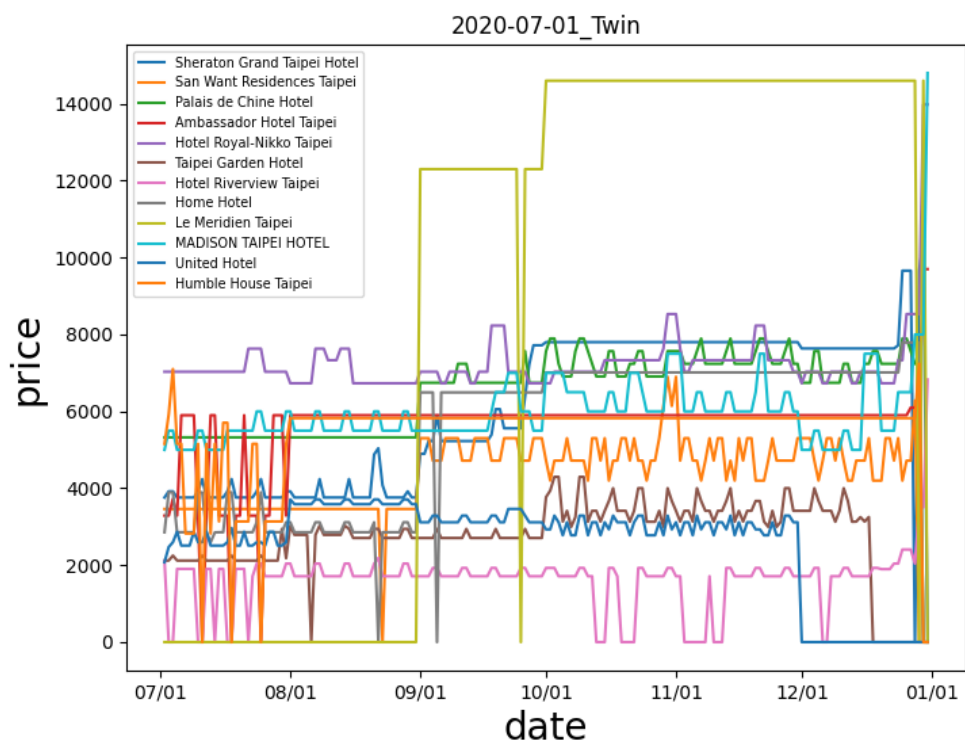
圖七、五星級飯店、King(單人房)價錢走勢圖

(前兩個是在台北車站、後兩個在101大樓)



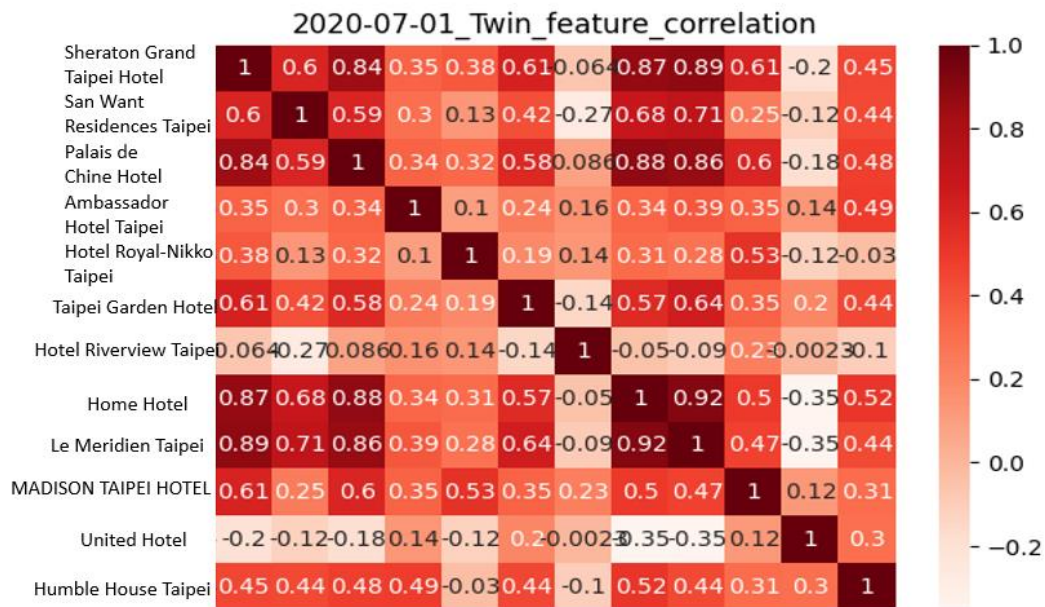
圖八、五星級不同飯店價錢走勢關係圖

(前兩個是在台北車站、後兩個在101大樓)



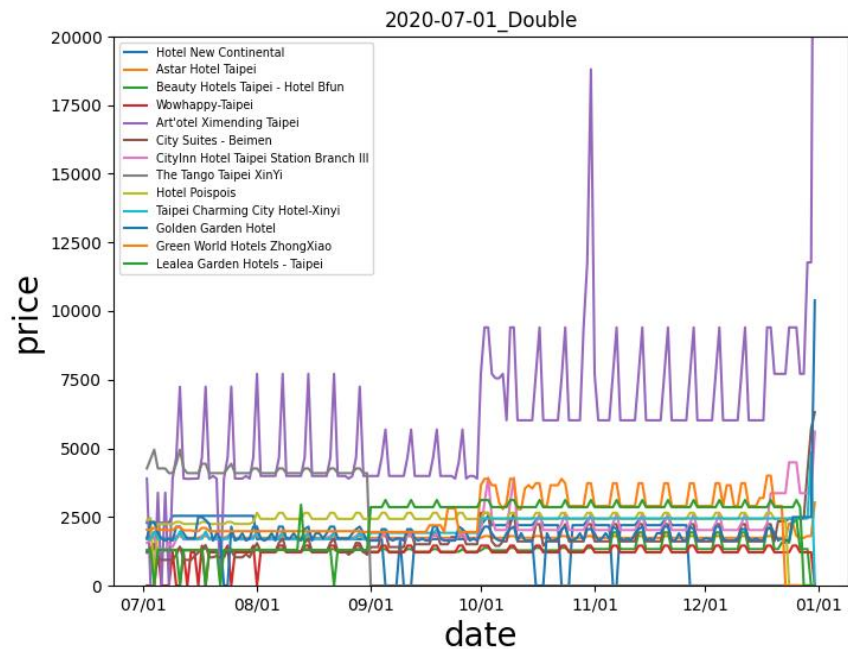
圖九、四星級飯店、Twin(單床雙人房) 價錢走勢圖

(前七個是在台北車站、後五個在101大樓)



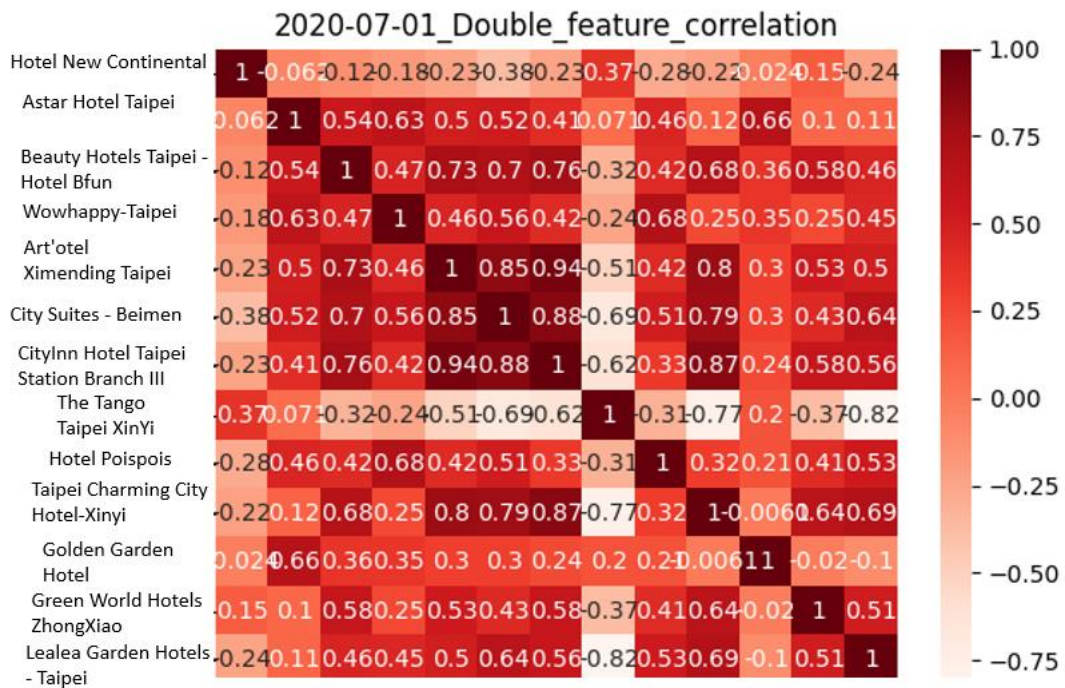
圖十、四星級不同飯店價錢走勢關係圖

(前七個是在台北車站、後五個在101大樓)



圖十一、三星級飯店、Double(雙床雙人房)價錢走勢圖

(前兩個是在台北車站、後三個在101大樓)



圖十二、三星級不同飯店價錢走勢關係圖

(前七個是在台北車站、後六個在101大樓)

(二)、觀察與發現

首先，根據圖七、圖八(07/01到07/01)，除了藍色的線在07/01時搜尋不到01/01後的資料，其餘三條線大部分的走勢大致相同，相關係數為0.31、0.32、0.56。值得注意的是綠線、紅線兩條走勢更為相近(如09/01到01/01間幾乎同時調整價錢)，說明相同地區關聯程度較大。再來是圖九、圖十(07/01到01/01)，發現大多數飯店的價錢波動也相當那麼明顯有週期性的變化，但是整體而言關聯性較五星級大，可是與地區較沒有太大關係。最後看圖十一、圖十二(07/01到01/01)，價格波動走勢波動最大，關聯程度是三種星級最大的，推估應是每間飯店價錢非常相近，為了吸引顧客，必須隨時依照市場、競爭對手調價。

陸、參考資料及其他

- [1] OTA學：時代趨勢下的旅宿產業思維(<https://solomo.xinmedia.com/globaltourismvision/141754>)
- [2] Expedia幫旅宿業帶進60億營收的工具大揭密(<https://www.bnext.com.tw/article/55536/explore-%2719-rev+>)
- [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-hot-el-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] 飯店房型怎麼分? (<https://www.xinmedia.com/article/90599>)
- [11] 如何避免爬蟲頻繁 requests server 導致被封鎖 IP (<https://python-learnnotebook.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.gtwang.org/programming/python-threading-multithreaded-programming-tutorial>)

專題工作進度

- [專題工作進度](#)
- [研究發現](#)
- [程式碼](#)
- [規劃](#)
- [Features for LSTM model](#)
- [Data processing](#)
- [Power Consumption Prediction using LSTM from Keras](#)
- [QA List for power prediction project](#)
- [Note about defining and training 2 layer lstm model in Keras](#)
- [作品說明書](#)
- [交流海報規劃](#)
- [觀光旅館營運統計](#)
- [旅遊業振興、補助事件](#)
- [Alex Feature](#)
- [工作進度](#)
- [下次進度:](#)
 - [Reference](#)

[研究發現](#)

[程式碼](#)

[規劃](#)

[Features for LSTM model](#)

[Data processing](#)

[Power Consumption Prediction using LSTM from Keras](#)

[QA List for power prediction project](#)

[Note about defining and training 2 layer lstm model in Keras](#)

[作品說明書](#)

[交流海報規劃](#)

[觀光旅館營運統計](#)

[旅遊業振興、補助事件](#)

[Alex Feature](#)

[工作進度](#)

9/10:

- *Action Item:*
 - ☒ *Task 1:: 閱讀資料 - [DeFi 导论](#)*
 - *Expected Outcome:*ppt
 - *Deadline:*1hr
 - *Status:*(Done)
 - ☒ *Task 2:: 旅館房價數據分析*
 - *Expected Outcome:嘗試一些統計模型來分析旅館房價資料 時間序列迴歸法、ARIMA 分析、分解法、指數平滑法
[scikit-learn](#)
 - *Deadline:*2-3hr
 - *Status:*(On-going)目前選定採用curve fitting 及 machine learning 方式推估出預測曲線
- *Reference:*
 - 1.[時間序列分析總體經濟與財務金融之應用](#)
 - 2.[MATLAB在時間序列建模預測及程序代碼](#)
 - 3.[Moving Averages](#)
 - 4.[Pandas索引的運用](#)
 - 5.[scipy.optimize.curve_fit](#)
 - 6.[Python !\[\]\(a22ba4e13c745edbf29e51af246c4c12_img.jpg\) Curve Fit with Step Test Data](#)

9/24:

- *Action Item:*
 - ☒ *Task 1:: 旅館房價數據分析*
 - *Expected Outcome:curve fitting
(<https://sites.google.com/site/ai20ntu/slides>)
 - *Deadline:*2-3hr
 - *Status:*(On-going)運用curve fitting
- *Homework:*

閱讀 中本聰白皮書
- *Reference:*
 - 1.[Using scipy for data fitting](#)
 - 2.[【python資料探勘課程】十四.Scipy呼叫curve_fit實現曲線擬合](#)
 - 3.[numpy.polyfit 和 numpy.poly1d 使用](#)
 - 4.[tensorflow 曲線擬合](#)
 - 5.[使用神经网络拟合曲线（MATLAB/Python）](#)
 - 6.[python最小二乘和神经网络拟合曲线比较](#)
 - 7.[BP神经网络的非线性曲线拟合和预测（未完）](#)

9/10:

- *Action Item:*
 - ☒ *Task 1:: 閱讀資料 - [DeFi 导论](#)*
 - *Expected Outcome:*ppt
 - *Deadline:*1hr
 - *Status:*(Done)
 - ☒ *Task 2:: 旅館房價數據分析*
 - **Expected Outcome:*嘗試一些統計模型來分析旅館房價資料 時間序列迴歸法、ARIMA 分析、分解法、指數平滑法
[scikit-learnAAAAI](#)
 - *Deadline:*2-3hr
 - *Status:*(On-going)目前選定採用curve fitting 及 machine learning 方式推估出預測曲線
- *Reference:*
 - 1.[時間序列分析總體經濟與財務金融之應用](#)
 - 2.[MATLAB在時間序列建模預測及程序代碼](#)
 - 3.[Moving Averages](#)
 - 4.[Pandas索引的運用](#)
 - 5.[scipy.optimize.curve_fit](#)
 - 6.[Python !\[\]\(83eb2aa26b610eb6a9dca7cf4702d681_img.jpg\) Curve Fit with Step Test Data](#)

9/24:

- *Action Item:*
 - ☒ *Task 1:: 旅館房價數據分析*
 - **Expected Outcome:*尋找利用AI進行房價預測的方式
 - *Deadline:*2-3hr
 - *Status:*(On-going)
- *Reference:*
 - 1.[使用Tensorflow建立RNN實戰：股市預測](#)
 - 2.[神經網絡預測股票市場](#)
 - 3.[實戰五：手把手教你用TensorFlow進行房價預測](#)

10/15:

- *Action Item:*
 - ☒ *Task 1:: 處理爬蟲程式的error*
 - **error:*

```
InsecureRequestWarning: Unverified HTTPS request
is being made to host 'www.hotels.com'.
Adding certificate verification is strongly advised.
```

- See: <https://urllib3.readthedocs.io/en/latest/advanced-usage.html#ssl-warnings>
- `InsecureRequestWarning`,

- *Deadline*:2-3hr
- *Status*:(On-going)

- *Reference*:

- 1.[python遷入信任憑證\(Adding certificate verification\)](#)
- 2.[SSL: CERTIFICATE VERIFY FAILED](#)

10/21:

- *Action Item*:

- ☒ *Task 1*:: 處理爬蟲程式的error
 - *solve:加 roomtypes += soup.find_all('li' ,class_ = 'room cont clea rfix mr-cell')
 - *Deadline*:2-3hr
 - *Status*:(done)
- ☒ *Task 2*:: 幫程式下註解

10/29:

- *Action Item*:

- ☒ *Task 1*:: 學習LSTM預測分析
 - *Status*:(On-going)

- *Reference*:

- 1.[Google Colab Free GPU Tutorial](<https://medium.com/deep-learning-turkey/google-colab-free-gpu-tutorial-e113627b9f5d>)
- 2.[LSTM](#)
- 3.[How to predict Bitcoin and Ethereum price with RNN-LSTM in Keras](#)
- 4.[\[Keras\] 利用Keras建構LSTM模型，以Stock Prediction 為例 1](#)

11/12:

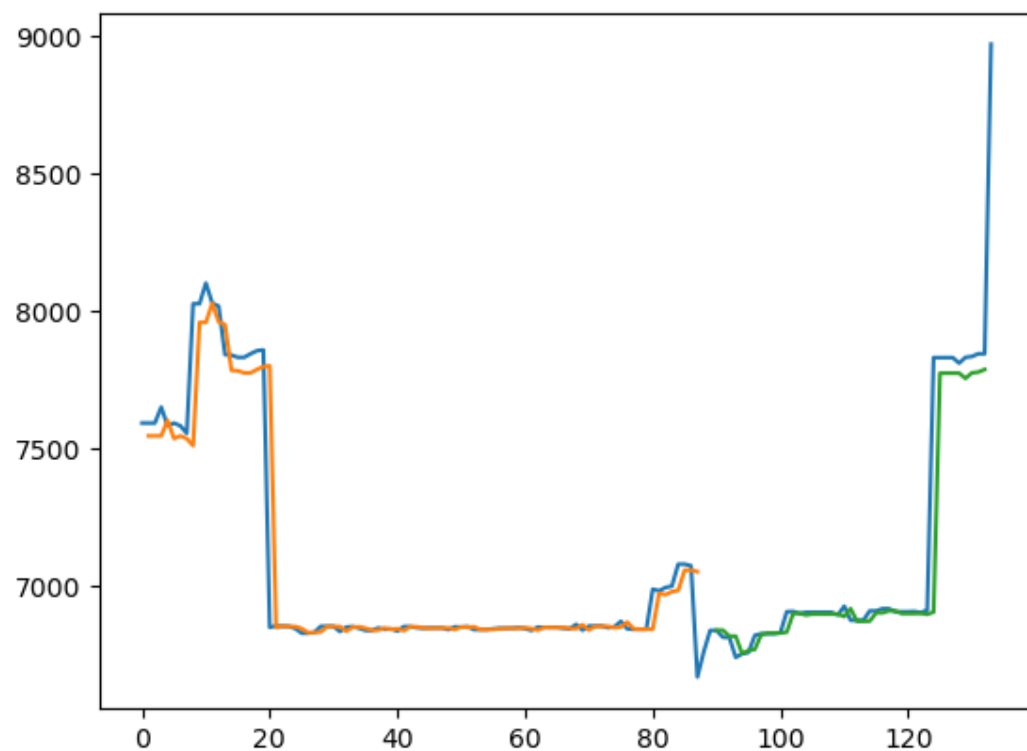
- *Action Item*:

- ☒ *Task 1*:: 進行LSTM預測分析
 - *Status*:(On-going)

- *Reference:*
 1. [銷售量預測 - LSTM 的另一個應用](#)
 2. [銷售量預測\(2\) - 『時間序列分析』技巧篇](#)

11/13:

- *Action Item:*
 - ☒ *Task 1::* 進行LSTM預測分析(初步結果)
 - *Status:*(On-going)

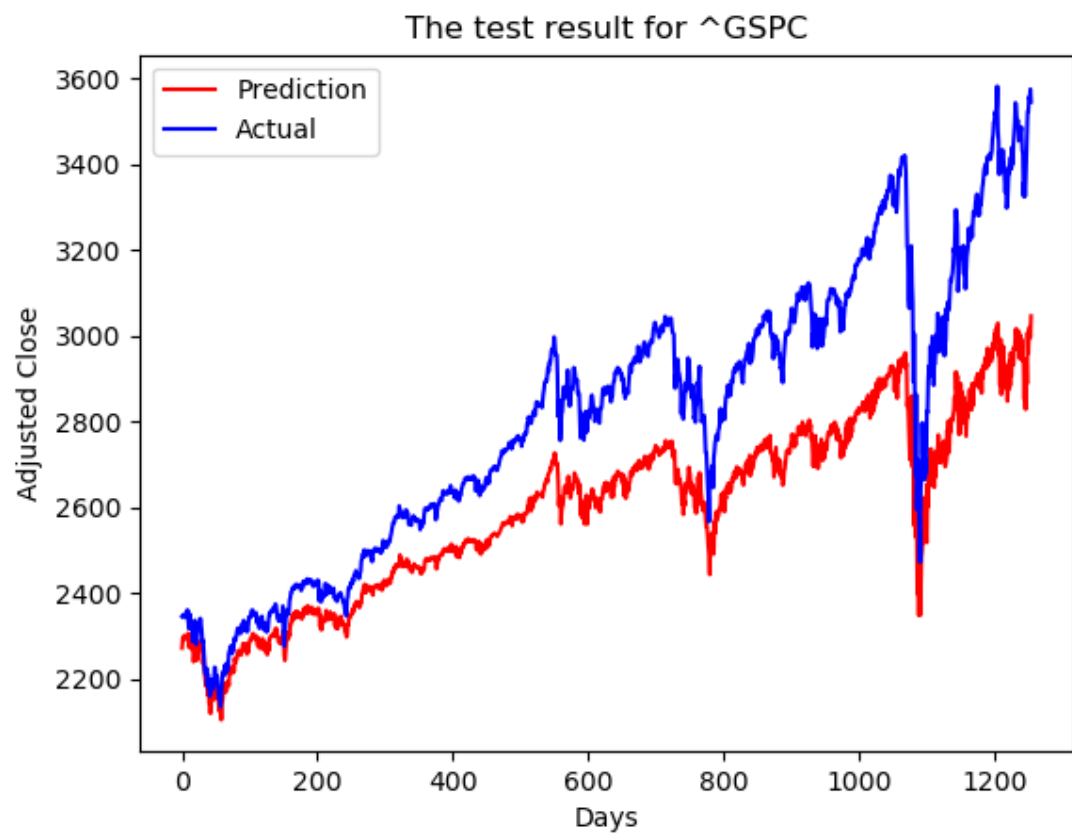
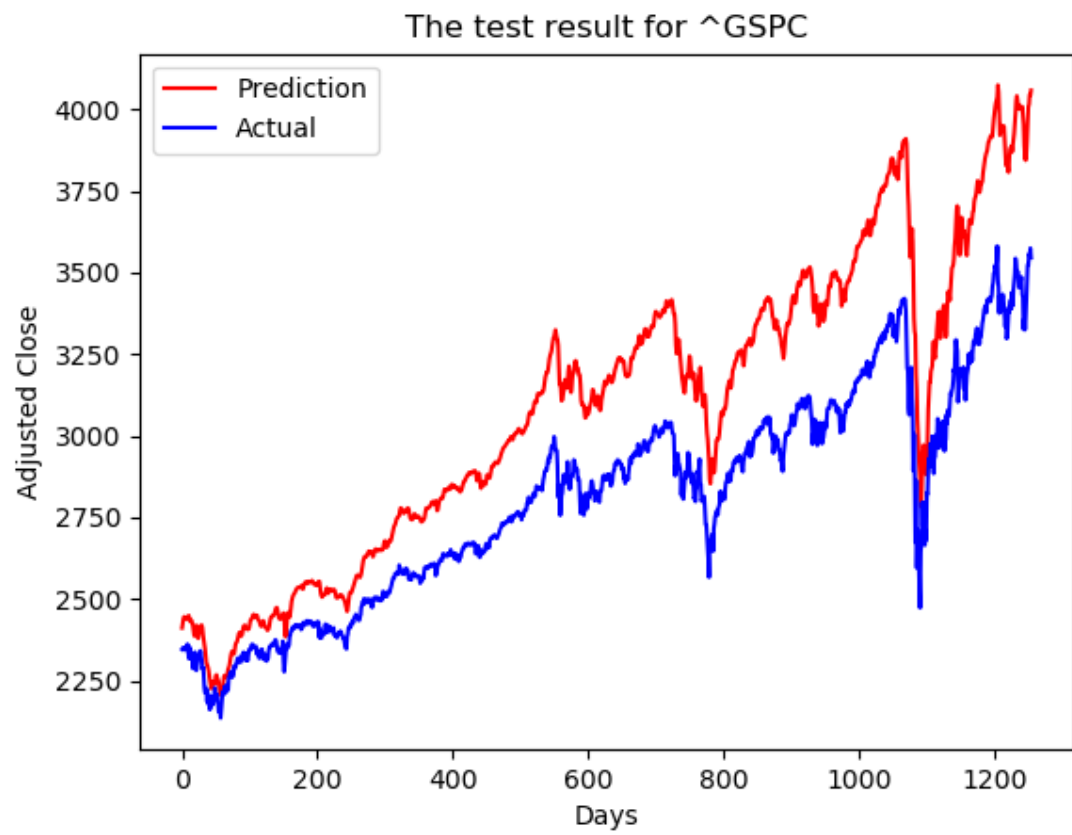


- *Reference:*
 1. [Keras學習筆記四：序貫模型（Sequential）](#)
 2. [Keras實戰：基於LSTM的股價預測方法\(吐血推薦，新手入門必看~\)](#)

11/14:

- *Action Item:*
 - ☒ *Task 1::* 進行LSTM預測分析(初步結果)

- *Status:(On-going)*



- *Reference:*
 - 1.[dataframe.as_matrix\(\)的作用](#)

11/19:

- *Action Item:*
 - ☒ *Task 1::* 將範例data改為自己的資料，進行LSTM預測分析
 - *Status:*(On-going)
- *Reference:*
 - 1.[遞歸神經網路（RNN）和長短期記憶模型（LSTM）的運作原理](#)
 - 2.[How to predict new samples with your Keras model?](#)
 - 3.[Excel 的規格及限制](#)
 - 4.[chrome driver session not created](#)

12/3:

- *Action Item:*
 - ☒ *Task 1::* 將範例data改為自己的資料，進行LSTM預測分析(hotel_lstm_1.py)
 - *Status:*(On-going)
- *Reference:*
 - 1.[Keras指南](#)

12/10:

- *Action Item:*
 - ☒ *Task 1::* 將範例data改為自己的資料，進行LSTM預測分析(hotel_lstm_1.py)
 - *Status:*(Done)
 - ☒ *Task 2::* 將data的第一欄索引值改成checkin date
 - *Status:*(Done)
- *Reference:*
 - 1.[Pandas資料的取得與篩選](#)
 - 2.[Pandas索引的運用](#)
 - 3.[dataset.columns](#)

12/17:

- *Action Item:*
 - ☒ *Task 1::* 選擇features，並實作LSTM
 - *Status:*(On-going)

delta	60 days ago	59 days ago	...	1 day ago	today(predict)
price	7000	8000	...	7000	7000

- *Reference:*

12/24:

- *Action Item:*
 - ☒ *Task 1::* 選擇features，並實作LSTM
 - *Status:*(On-going)

- *Reference:*
 - 1.[Python endswith\(\)方法](#)
 - 2.[Google試算表嵌入](#)
 - 3.[Why Feature Correlation Matters ... A Lot!](#)

12/25:

- *Action Item:*
 - ☒ *Task 1::* 建造holiday的dict

```
vacation = { 'New Year' :[ '2021-01-01' , '2021-01-02' , '2021-01-03' ], 'Spring Festivals' :[ '2021-02-10' , '2021-02-11' , '2021-02-12' , '2021-02-13' , '2021-02-14' , '2021-02-15' , '2021-02-16' ], '228' :[ '2021-02-27' , '2021-02-28' , '2021-03-01' ], 'Ching Ming Festival' :[ '2021-04-02' , '2021-04-03' , '2021-04-04' , '2021-04-05' ], 'Dragon Boat Festival' :[ '2020-06-25' , '2020-06-26' , '2020-06-27' , '2020-06-28' , '2021-06-12' , '2021-06-13' , '2021-06-14' ], 'Moon Festival' :[ '2020-10-01' , '2020-10-02' , '2020-10-03' , '2020-10-04' , '2021-09-18' , '2021-09-19' , '2021-09-20' , '2021-09-21' ], 'National Day' :[ '2020-10-09' , '2020-10-10' , '2020-10-11' , '2021-10-09' , '2021-10-10' , '2021-10-11' ],}
```

-*Status:*(Done)

- *Reference:*
 - 1.[python module - chinese holiday](#)
 - 2.[python module - holidays](#)
 - 3.[matplotlib.pyplot，如何更改圖例大小](#)

12/31:

- *Action Item:*

- ☒ *Task 1::* 製作交流海報
- *Reference:*
 - 1.[matplotlib命令與格式：圖例legend語法及設置](#)
 - 2.[如何在 Matplotlib 中設定軸的範圍](#)

1/7:

- *Action Item:*
 - ☒ *Task 1::* 製作交流海報
- *Reference:*
 - 1.[檢查字串中是否包含指定的子字串](#)
 - 2.[飯店房型怎麼分](#)

1/14:

- *Action Item:*
 - ☒ *Task 1::* 製作交流海報(針對同星級不同飯店相同房型相同搜尋日作圖、分析)
- *Reference:*
 - 1.[model.evaluate \(\)](#)
 - 2.[Keras model.fit\(\)参数详解](#)
 - 3.[numpy.ndarray 在求mean，max，min的时候如何忽略跳过nan值，使用np.nanmean, np.nanmax](#)
 - 4.[布林通道](#)
 - 5.[Python pandas.DataFrame.interpolate函数方法的使用](#)
 - 6.[LSTM introduction](#)
 - 7.[How Do Online Travel Agencies Work?](#)

下次進度:

-
- ☐ *Task 1::* 將範例data改為自己的資料，進行LSTM預測分析(hotel_lstm_2.py(在相同日期搜尋不同日期下相同房間之價格趨勢))
-Status:(On-going)
 - ☐ *Task 2::* 進行LSTM預測分析，使用features (hotel_lstm_timedelta(train).py)
-Status:(On-going)
 - ☐ *Task 3::* 尋找Features correlation，在選用適當feature
-Status:(On-going)

- Task 4:: 製作交流海報(針對同星級不同飯店相同房型相同搜尋日作圖、分析-4star、3star)
-Status:(On-going)

```

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)#處理空房間問題

```

```

67
68 print(price_total)
69 startdate = (datetime.datetime.now()+datetime.timedelta(days=1))
70 stardate = str(startdate.date())
71 for i in database:
72     print(len(price_total[i]))
73     df = pd.read_csv(r'C:/Users/USER/Desktop/資訊/hotels project/data/信義區
74 /Home Hotel(4.5星級)/'+i+'.csv',encoding='big5')#讀檔
75     df[today]=price_total[i]
76     print(df)
77     df.to_csv(r'C:/Users/USER/Desktop/資訊/hotels project/data/信義區
78 /Home Hotel(4.5星級)/'+i+'.csv',encoding='big5',index=False)#存檔

```

附圖一、程式碼

Analysis of Pricing Strategy of Online Travel Agent

Heng-An Cheng(鄭恆安)

The Affiliated Senior High School of National Taiwan Normal University

Abstract

This project aims to analyze the pricing strategy of a hotel based on the data collected from the online travel agent (OTA) platform. First, we develop a python-based program to automatically crawl the room rates everyday for building the database. Next, we use data visualization to identify key factors that affect the room rates. Then, we choose the key factors as inputs of a long short-term memory (LSTM) model and use the data collected from a hotel to train the LSTM model. Finally, we utilize the LSTM model to predict the room rates for a special room type between the reservation date and the target check-in date based on the historical room rates observed over the past 30 days before the reservation date.

Motivation & Purpose

OTAs, such as Hotels.com, Agoda, Booking.com, generally work on two models to make profit. That is Merchant Model and Agency Model. In order to make profit, OTAs adjust room rate using machine learning technology, according to the conditions of each market, and then obtain future market supply and demand. Thus, we began to find factors causing the price up and down and use machine learning to predict future price trend.

Introduction:

OTA

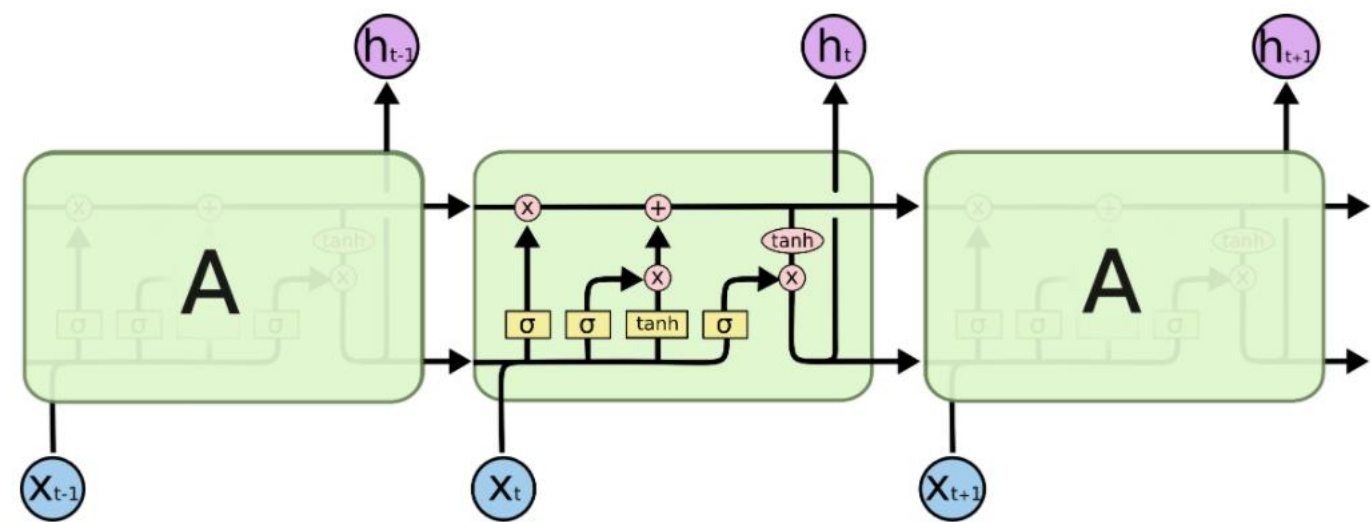
“OTA” stands for Online Travel Agency, which is a travel agency whose primary presence is on digital channels. Consumers can use a website and/or mobile device to search and book travel -- all without the traditional “gatekeeper” travel agent. OTAs connect to the full breadth of travel providers, giving travelers access to all of the inventory that they may want for their next trip.

Web Crawl

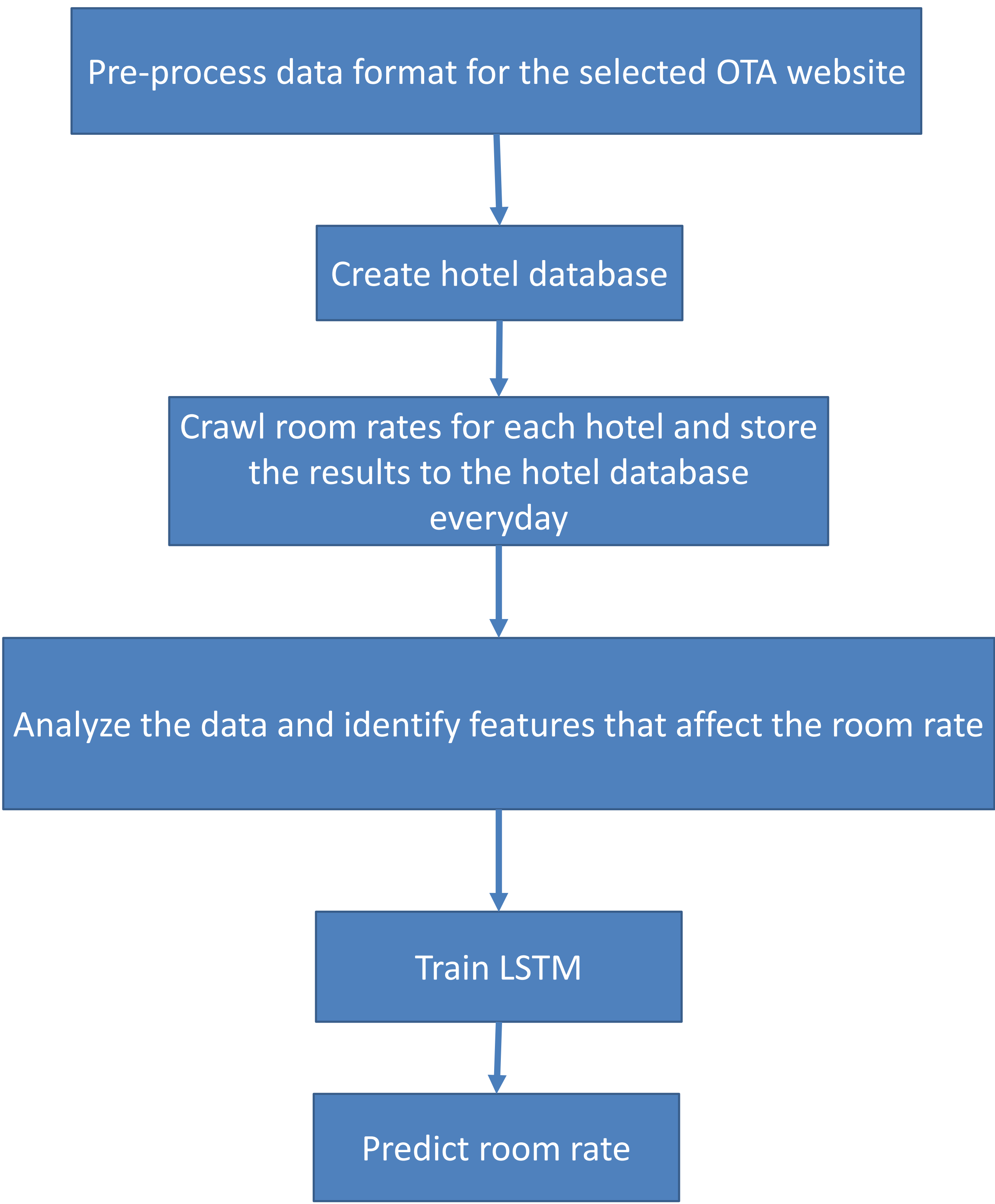
Crawl is a term describing a bot, script, or software program that visits a web page and grabs content and links from it. Once completed, the program visits the next link in the list or a link obtained from the web page it had recently visited.

LSTM

Long Short Term Memory networks – usually just called “LSTMs” – are a special kind of RNN, capable of learning long-term dependencies. They work tremendously well on a large variety of problems such as handwriting recognition, image classification, and price prediction.



Flow Chart(process):



Process:

Build room rate database

Since there isn't any opensource room rate data online, we have to build our own database. To meet our goal, we use python requests module to crawl the web page from hotels.com. After that, we use BeautifulSoup module to parse the web page and choose the room rate of different room types and hotels every day.

Data visualization analysis

According to previous researches, we find out some main factors that influence room rate. Therefore, we plot the data on charts to make it clear to find the correlation between each factor and the room rate. For example, through Fig1, we can discover the lowest price doesn't always occur on the earliest date you search, which means the interval may be one of factors.

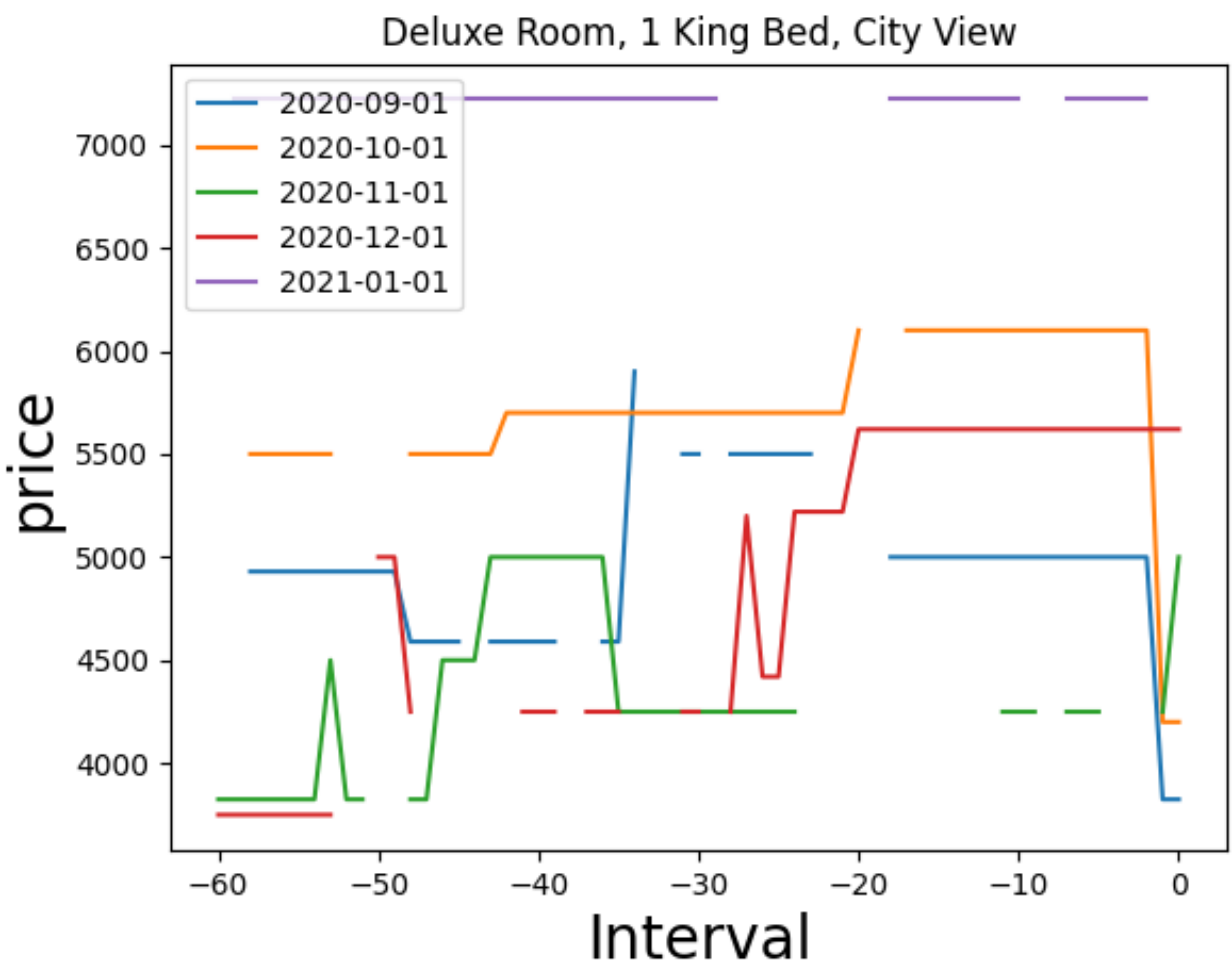


Fig 1. Different check-in date, the same interval between search date and check-in date

Train LSTM model

- 1.Data preprocessing : Because there is some missing data in our database, so we use interpolation to correct database.
- 2.Create features: we find out several factors through data visualization analysis. Check-in date(year, month, day, day of week, holiday), room type, interval between search date and check-in date, and the room rate searched on the search date.
- 3.Create train data : First, we normalize our data by using min max normalization. Then, we use previous 30 days data to predict tomorrow room rate.
- 4.Build model : Fig 2
5. Test model : Evaluate the mean square error .

6. Predict : Fig3

Model: "sequential_1"

Layer (type)	Output Shape	Param #
lstm_2 (LSTM)	(None, 30, 128)	70144
dropout_2 (Dropout)	(None, 30, 128)	0
lstm_3 (LSTM)	(None, 128)	131584
dropout_3 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 32)	4128
dense_3 (Dense)	(None, 1)	33
Total params: 205,889		
Trainable params: 205,889		
Non-trainable params: 0		

Fig 2. model summary

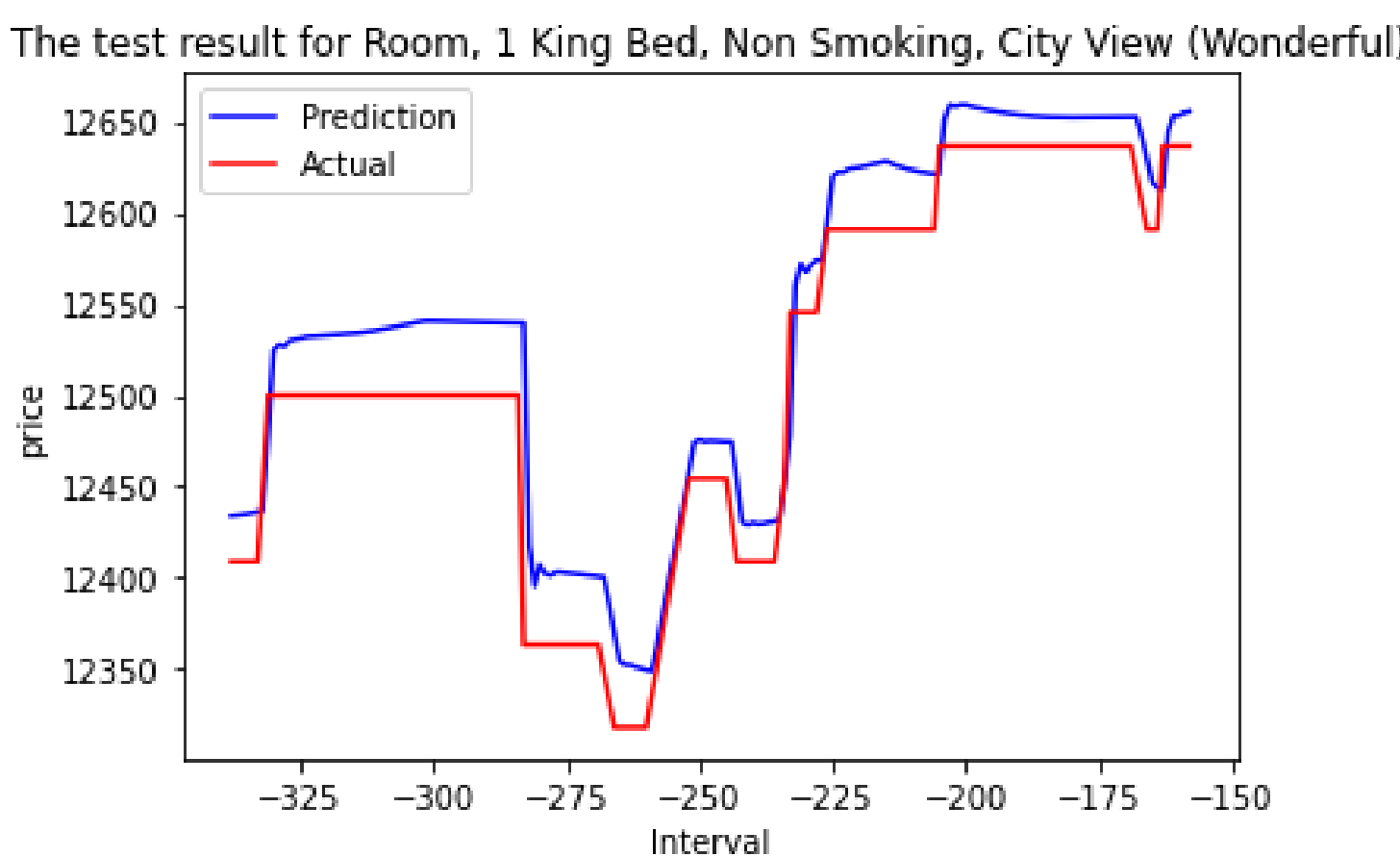


Fig 3. prediction and actual room rate