

中國文化大學商學院資訊管理學系  
碩士論文

Master's Thesis  
Department of Information Management  
College of Business  
Chinese Culture University

結合自然語言處理與情緒分析技術  
之 SWOT 分析－以旅遊電商  
KKday 為例

Applying Natural Language Processing and Sentiment  
Analysis on SWOT analysis - A case study of  
Travel Ecommerce Platform KKday

指導教授：杜淑芬教授  
Advisor: Professor Shu-Fen Tu

研 究 生：呂尉慈  
Graduate Student: Yu-Tzu Lu

中華民國 109 年 6 月  
June, 2020

中國文化大學  
研究所碩士班  
論文口試審定委員會

本校資訊管理學系碩士班呂尉慈君

所提論文結合自然語言處理與情緒分析技  
術之 SWOT 分析－以旅遊電商  
KKday 為例

合於碩士資格水準，業經本委員會評審認可。

口試委員：

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

指導教授：

\_\_\_\_\_

研究所所長：

\_\_\_\_\_

中 華 民 國 1 0 9 年 6 月

論文名稱：結合自然語言處理與情緒分析技術之 總頁數：64

SWOT分析-以旅遊電商KKday為例

校(院)所組別：中國文化大學商學院資訊管理學系碩士班

畢業時間及提要別：108學年度第2學期碩士學位論文提要

研究生：呂尉慈

指導教授：杜淑芬

論文提要內容：

本研究以台灣目的地旅遊體驗新創公司 KKday 為主要研究對象，利用 RE-SWOT 工具進行 KKday 的 SWOT 分析。RE-SWOT 工具以線上使用者評論建置 SWOT 分析矩陣，本研究採用世界最大的旅遊評論網站 TripAdvisor 的評論，而 RE-SWOT 工具會利用到評論附帶的星號評等，這個星號評等是由使用者主動提供的，然而 TripAdvisor 的評論並沒有使用者評分，本研究提出以情緒分數來替代 RE-SWOT 工具中所需之使用者評分，情緒分數則是使用 Python VADER 情緒分析套件產生。

根據研究結果顯示，KKday 具備的優勢有導遊服務好、便宜且方便使用的景觀台門票、線上信用卡支付方便、使用二維條碼的票券不需現場領票、多選擇的旅遊活動、電話卡的販售領取方便以及軌道自行車和纜車的價格優惠又有趣，而與日月潭有關的產品和服務為待改善之處；此外，Klook 在各式活動提供的優惠折扣較受消費者歡迎，是 KKday 需要注意的威脅因子，以防止 KKday 使用者轉移至 Klook。最後，本研究的方法可以延伸擴展至不同領域的企業，以了解企業和其競爭對手的相關優劣勢。

關鍵字：目的地旅遊體驗平台(local in-destination tours and activities platform)，情緒分析(sentiment analysis)，自然語言處理(nature language processing)，RE-SWOT 分析(requirements engineering swot)，KKday

Applying Natural Language Processing and Sentiment

Analysis on SWOT analysis - A case study of

Travel Ecommerce Platform KKday

Student: Yu-Tzu Lu

Advisor: Shu-Fen Tu

## Chinese Culture University

### ABSTRACT

This study utilizes RE-SWOT tool to perform a SWOT analysis on KKday, which is a Taiwanese local in-destination tours and activities platform. The RE-SWOT tool retrieved online user reviews and ratings to construct SWOT matrix automatically. In this study, the user reviews are retrieved from TripAdvisor because TripAdvisor is the world's largest travel information platform. However, TripAdvisor is lack of user ratings, which are necessary for calculating Feature Performance Score (FPS). This research proposed to make sentiment analysis on user reviews to generate user ratings, and the python package, VADER, was adopted to perform sentiment analysis.

According to the research results, the advantages of KKday include: good tour guide, cheap and convenient tickets of observation deck, convenient online payment by credit card, QR code tickets instead of on-site ticket collection, various tourist activities, easy to collect phone cards, and cheap and interesting rail bikes and cable cars. The activities related to Sun Moon Lake needs to be improved. Besides, the discounts on various activities offered by Klook are more popular with consumers, and hence are a threat that KKday needs to pay attention to prevent KKday customers from transferring to Klook. The proposed method can be extended to other business to help managers to understand the relative advantages and disadvantages of the company and its competitors.

Key Words : Local In-destination Tours and Activities Platform, Sentiment Analysis,

Nature Language Processing, Requirements Engineering SWOT, KKday



## 誌 謝 辭

想當初剛回台灣，懵懂的進入文化大學資訊管理學系就讀，一轉眼五年過去了，在文化大學求學生涯，就在這本論文寫作中告一段落。這一年的碩士生活相當充實，感謝有著許多人的協助、給予意見和鼓勵，讓我一關一關順利過。

誠摯的感謝我的碩士論文指導教授，杜淑芬博士，提供我撰寫論文的指引方針，督促我論文進度的進程，並且在每個章節提出撰寫上的修正意見，讓我能逐步完成碩士論文，若不是教授對於我的督促，實在很難想像能在短短的幾個月內，發揮無限的潛力完成此篇碩士論文。感謝校內口試委員盧惠莉博士以及校外口試委員許慶昇博士於百忙之中，撥冗審查本論文，並針對論文提出寶貴的意見，讓本研究的論述更臻完備。並且感謝所有資訊管理學系的所有師長們的教導，有些無法一一致謝，但心存感激。

感謝我的同學在我撰寫論文時，提供程式撰寫以及其他課程學習上的協助與討論。最後，感謝家人的支持，讓我能無後顧之憂的完成學業。謹以本論文獻給所有關心我、愛護我、提攜我的人。

# 內 容 目 錄

中文摘要	iii
英文摘要	iv
誌謝辭	vi
內容目錄	vii
表目錄	ix
圖目錄	x
第一章 緒論	1
第一節 研究背景	1
第二節 研究動機	3
第三節 研究目的	7
第二章 文獻探討	8
第一節 目的地旅遊體驗平台	8
第二節 情緒分析	12
第三節 自然語言處理	14
第四節 RE-SWOT 分析	17
第三章 研究方法	24
第一節 研究架構	24
第四章 研究結果	33
第一節 數據來源	33
第二節 特徵詞組	35
第三節 特徵 FPS	31
第四節 RE-SWOT 分析	43
第五節 生成需求	46
第五章 結論	48
第一節 研究意涵	48

第二節	研究限制	50
第三節	未來展望	50
參考文獻		52





## 表 目 錄

表 1-1	SWOT 分析相關研究領域 . . . . .	4
表 2-1	KKday 的大事紀 . . . . .	9
表 2-2	Klook 的大事紀 . . . . .	10
表 2-3	SWOT 分析矩陣 . . . . .	18
表 2-4	RE-SWOT 分析矩陣 . . . . .	21
表 2-5	RE-SWOT 分析矩陣範例 . . . . .	22
表 3-1	詞性列表 . . . . .	28
表 3-2	RE-SWOT 分析矩陣範例 . . . . .	32
表 4-1	KKday 及 Klook 評論產生詞組 . . . . .	34
表 4-2	選用之特徵詞組 . . . . .	35
表 4-3	KKday 及 Klook 特徵之 FPS ( $\sigma=0.1$ ) . . . . .	36
表 4-4	KKday 及 Klook 特徵之 FPS ( $\sigma=0.2$ ) . . . . .	38
表 4-5	KKday 及 Klook 特徵之 FPS ( $\sigma=0.3$ ) . . . . .	39
表 4-6	KKday 及 Klook 特徵之 FPS ( $\sigma=0.4$ ) . . . . .	41
表 4-7	KKday 之 RE-SWOT 分析結果( $\sigma=0.1$ ) . . . . .	43
表 4-8	KKday 之 RE-SWOT 分析結果( $\sigma=0.2$ ) . . . . .	44
表 4-9	KKday 之 RE-SWOT 分析結果( $\sigma=0.3$ ) . . . . .	44
表 4-10	KKday 之 RE-SWOT 分析結果( $\sigma=0.4$ ) . . . . .	45

## 圖 目 錄

圖 1-1	傳統旅遊產業鏈 . . . . .	2
圖 1-2	旅遊產業鏈轉變 . . . . .	2
圖 2-1	RE-SWOT 流程圖 . . . . .	18
圖 3-1	研究流程圖 . . . . .	25



# 第一章 緒論

## 第一節 研究背景

觀光具有悠久的歷史，依據聯合國世界觀光組織(World Tourism Organization, UNWTO)的定義，觀光指的是人在其常駐地以外的地區從事旅遊活動，且停留持續不超過一年。旅遊活動最大的特色是能引起人際間的交流，並使環境及空間產生複雜的轉換關係，這樣的特色使人們旅遊時需要求助於旅行社的協助，最早的旅行社創始人是英國人托馬斯·庫克(Thomas Cook)，他為協助顧客旅途上所需的事物而成立通濟隆旅遊公司。傳統的旅行社產品型態可以依據所包含的內容分為：包價旅遊、組合旅遊和單向旅遊。包價旅遊通常是以摺包行程和團體旅遊的形式呈現，由旅客預先向旅行社支付服務內容的費用，旅行社提供的產品是將各類旅遊產品中的要素(食、住、行、遊、購、娛等)和其提供的服務和附加價值(諮詢、導游、手續、保險等)組合起來。組合旅遊主要的形式是在旅遊線路，產品的提供者是旅遊目的地的旅行社，他們根據對旅遊客源的市場需求做調查和分析，規劃出一批固定的旅遊線路，通過各客源地旅行社的推廣、宣傳、銷售，遊客自行安排按時到達旅遊目的地的方式，再由目的地旅行社將遊客集中起來組成團體活動。單項旅遊產品是旅行社根據旅客的具體需求而提供的各種單項付費服務。遊客的需求多樣性決定了旅行社單項服務的可能性和廣泛性(韓勇，2006)。

傳統的旅遊產業鏈是由旅遊產品供應商、旅遊產品批發商、旅遊產品零售商和最終消費者串接組成如圖 1-1 所示。旅遊產品供應商泛指食、住、行、遊、購、娛等旅遊要素供應商。連接價值鏈的中間商像是旅行社這種旅遊產品批發商及零售商提供服務，最終才到消費者。

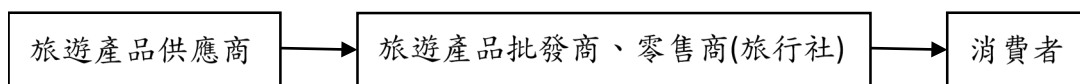


圖 1-1 傳統旅遊產業鏈(亮，2016)

近年來網路科技的進步，在旅遊產業鏈也發生變化。傳統的產業鏈中旅遊產品供應商難以直接接觸旅客，需透過旅行社才能與旅客互動。加上在自由旅行及電子商務盛行的趨勢下，興起了旅遊電商平台，且強調目的地體驗行程，此篇論文將其稱為目的地旅遊體驗平台(local in-destination tours and activities platform)。此平台是線上旅客及線下旅遊產品供應商的結合，去中間化使供應商可直接與消費者互動，可以提供更透明化與完整的服務資訊，也使消費者有更多元化的選擇如圖 1-2 所示。這樣的平台有比線下旅遊業者更優惠的價格，又著重於目的地主題式探索，進行優惠促銷，或是以使用者所屬地區的假期特色、文化背景為出發點設計優惠。亦有推薦在地的一日或多日遊體驗行程、租借網卡、交通或是在地美食。再加上線上的網路購物程序方便，可以快速網上選購，再透過電子票券就可以輕鬆使用。目的地旅遊體驗平台帶動了整個旅遊消費型態的改變，讓原本要從實體線下通路購買的旅客消費習慣轉往線上的網購模式，適合忙碌沒時間規劃旅程的上班族旅客、到陌生城市旅遊自由行旅客或是想要有比較更多行程的人能更輕鬆預訂及安排行程(何佩珊，2018a；KATI，2019；張文騰，2017；亮，2016)。

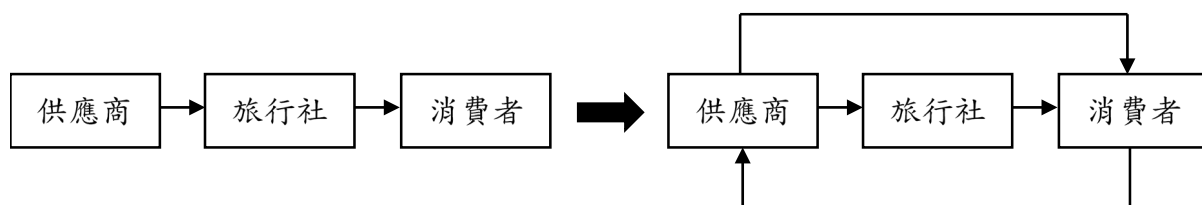


圖 1-2 旅遊產業鏈轉變(亮，2016)

全球各地區都有目的地旅遊體驗平台，此外，Booking.com、

Expedia 等 OTA 也積極佈局該領域。根據旅遊產業分析公司 Phocuswright 研究成果顯示，目的地行程、景點、消費等活動為成長最快的旅遊業務，市場規模預計將由 2016 年的美金 1,350 億成長到 2020 年的 1,830 億。其中三分之一市場將集中在亞洲，由此可知，OTA(Online Travel Agency)、目的地旅遊體驗平台未來可望掌握龐大商機。在百家爭鳴的市場中，如何成為一家新創脫穎而出、坐大市場便是關注的焦點(楊孟芯，2019)。

各地區的目的地旅遊體驗平台，像是早期 2009 年德國成立 GetYourGuide、2012 年美國成立 Peek，亞洲的目的地旅遊體驗行程預定平台還是非常分散的。直到 2014 年亞洲旅遊電商平台的興起，在台灣也成立了第一個專門提供旅遊體驗與行程旅遊電商平台 KKday，其創辦人陳明明熱愛自由行，並且具備線上旅遊產業的經驗近 20 年，看好旅遊電商產業的前景及發展性。KKday 提供旅人來自世界各地的多樣化在地體驗行程，讓旅行能更深入當地，在規劃上也更輕鬆、更自由，因此興起了「目的地旅遊」的浪潮。KKday 平台服務遍布全球超過 80 個國家和 500 個城市、提供的旅遊行程與體驗超過 20,000 種(KKday，2019)。平台以自由行為基礎，深入開發各種類型目的地旅遊體驗行程，並提供易於使用的預訂平台，使旅客可以彈性的組合體驗行程，安排出最合適的旅遊方式，這樣的模式也形成了近年自由行程的網購風潮。

## 第二節 研究動機

SWOT 分析是 20 世紀 80 年代初由美國舊金山大學的管理學教授海因茨·韋里克(Heinz Weihrich)提出，此種分析方法被廣泛運用於各個領域，並且能依據 SWOT 分析結果進行市場分析、經營策略、行銷計畫、產品推動、競爭者分析等計畫(Lien，2020)，如表 1-1 所示。MarketingTeacher(2017)的 SWOT 示例中說明，Google、

Nike、eBay、Walt Disney、Facebook 等使用 SWOT 分析，從中了解其優勢(Strengths)、劣勢(Weaknesses)、機會(Opportunities)和威脅(Threats)。除了企業廣泛使用 SWOT 分析之外，當今研究中各領域也都使用 SWOT 分析，Kamran et al. (2020)研究，巴基斯坦培育可再生能源部門的現狀和未來路線圖可行性的參考點，通過 SWOT 確定了對可再生能源行業可持續發展的社會、經濟和環境影響；Wang and Wang (2020)研究，使用 SWOT 分析方法，針對中國 COVID-19 流行病的預防和控制策略；Ivanov (2019)研究，針對對企業銷售管理活動，應用 SWOT 分析了解其優勢與劣勢及四種策略建議；Niranjanamurthy et al. (2019)研究，利用 SWOT 分析，分析區塊鏈的類型以及區塊鏈如何工作以及其優缺點；Aksu and Bayar (2019)研究，透過 SWOT 分析對土耳其健康旅遊發展提供意見。

表 1-1 SWOT 分析相關研究領域

研究	研究領域
MarketingTeacher(2017)	Google、Nike、eBay、Walt Disney、Facebook 等企業戰略。
Kamran et al. (2020)	對可再生能源行業可持續發展的社會、經濟和環境影響。
Wang and Wang (2020)	針對中國 COVID-19 流行病的預防和控制策略。
Ivanov (2019)	針對企業銷售管理活動。
Niranjanamurthy et al. (2019)	區塊鏈的類型以及區塊鏈如何工作以及其優缺點。
Aksu and Bayar (2019)	對土耳其健康旅遊發展提供意見。

社交媒體的發展從根本上改變了與旅遊有關的所有業者和消

費者的生產及消費資訊方式。使通過社交媒體共享的旅遊內容已成為影響力極大的資訊來源，無論是在業者聲譽還是在績效方面都對旅遊產生影響(Alaei, 2019)，因此學者開始運用線上評論來了解業者市場定位，例如 Zhao and Wang (2019)根據大量的飯店線上評論預測整體客戶滿意度；Qi et al. (2016)從線上評論中挖掘客戶需求及產品改善的角度；Ahani et al. (2019)通過線上評論進行的市場區隔，幫助 Spa 酒店更好地為行銷策略制定提供訊息。因此若能以線上使用者的評論為基礎，為企業的產品提供 SWOT 分析，將更有助於企業從使用者的觀點瞭解自身的競爭優勢，訂定有效的競爭策略。

Parente (2018)所提出的 RE-SWOT，即是以使用者評論為基礎，來產生 SWOT 分析矩陣的一種工具。RE-SWOT 以行動應用 (Mobile App)軟體為研究對象，擷取軟體在應用商店(app store)中的使用者評論及評等(rating)，建置行動應用軟體的 SWOT 分析 (SWOT Analysis)矩陣，以瞭解使用者對行動應用產品的需求。RE-SWOT 工具使用自然語言處理(Natural Language Processing, NLP)和可視化(Information Visualization, IV)技術，將公司自身產品的使用者評論轉換成內部能力(優勢與劣勢)，將主要競爭者產品的使用者評論轉換成外部因素(機會與威脅)，以了解自身產品的發展方向，像是 Dalpiaz and Parente (2019)用行動應用的線上評論，通過 RE-SWOT 得出三家行動應用公司間的競爭分析及其用戶的需求資源，以幫助公司根據用戶評論做出即時決策。

本研究希望採用 RE-SWOT 工具，針對 KKday 進行 SWOT 分析，並且以 KKday 的首要競爭者 Klook (Owler, 2019)做為 SWOT 外部因素的來源。Klook 與 KKday 在同一年成立，Klook 成立於香港，其服務涵蓋世界各地共 250 多個城市，提供 8 萬多種景點、特色活動、交通及美食預訂選擇。Klook 同樣擁有來自世界各地

的旅遊體驗師親自嚴選當地活動。嚴格篩選當地營運業者，與其直接合作繞過中間所有環節。RE-SWOT 工具需要擷取使用者針對公司產品的評論與評等，KKday 與 Klook 官網兩者皆有，因此若取用官網的評論與評等，可直接套用 RE-SWOT 工具進行 SWOT 分析，然而根據呂尉慈與杜淑芬(2020)的研究，發現 KKday 及 Klook 官網中，使用者給予的評論與評等皆極為正向，由於官網的後台管理控制權在公司，若取用官網的評論與評等，恐有失客觀。因此，選擇一個合適的使用者評論來源，是本研究情境中採用 RE-SWOT 工具需要考量的一個重要課題。而在 Parente (2018) 的論文中，其研究對象為行動應用軟體，而使用者的評論與評等是來自於行動應用商店，行動應用商店對於行動應用軟體開發公司而言，是個第三方平台，不易受到單一公司的操控，因此在 Parente 的研究情境中，並不需要考慮客觀性的問題。

TripAdvisor 是一個被廣泛使用的旅遊網站，使用者遍佈全世界各國，為全球旅遊者最大之旅遊平台，TripAdvisor 網站的線上評論因為較無人為操控的疑慮，也成為許多旅遊相關研究使用的資料來源(游綉雯，2015)，因此本研究將爬取 TripAdvisor 論壇(Forum)有提到 KKday 和 Klook 的使用者評論，而不採用官網的使用者評論。然而，TripAdvisor 論壇僅有使用者留言的文字訊息，並沒有使用者評等，以 TripAdvisor 論壇做為 RE-SWOT 工具的資料來源時，如何補上使用者評等，便成為一個待解決的問題。

當今的企業嚴重依賴數據，大多數數據是來自於電子郵件、聊天、社交媒體、調查、文章和文檔等非結構化文本。篩選大量非結構化文本數據既困難又費時。此外，它需要大量的專業知識和資源來分析所有這些文本。並且情緒分析對從業人員和研究人員也很有用，尤其是在社會學、市場營銷、廣告、心理學、經濟學和政治學等領域，這些領域非常依賴人機交互數據。因此，礙



於由於現今網絡上的數據量已達到幾乎無法進行手動處理的水平，情感分析作為一種檢查評論中語義關係和含義的自動化過程變得相當重要(Pandey, 2018)。由於進行 RE-SWOT 分析時需要使用評分，但是由於本研究取用 TripAdvisor 論壇中的評論沒有等級，因此我們將使用情緒分析方法針對評論去產生評論等級來代替 RE-SWOT 分析時所需要的使用者評分。

### 第三節 研究目的

本研究將以台灣新創公司 KKday 為主要研究對象，這種目的地旅遊體驗平台提供各地民眾目的地旅遊資訊及線上預訂目的地旅遊體驗行程等旅遊相關服務，在旅遊相關研究中，卻沒有特別針對 KKday 做競爭分析的相關研究。本研究會採用 RE-SWOT 工具，SWOT 分析的內部能力來自於 TripAdvisor 論壇中提到 KKday 的使用者評論，SWOT 分析的外部因素，則來自於 TripAdvisor 論壇中提到 Klook 的使用者評論。如前所述，TripAdvisor 論壇的留言缺少使用者評等，本研究提出以情緒分析的方法，來自動化產生每則評論的情緒分數，並且以情緒分數轉換成使用者評等，如此一來，便可使用 RE-SWOT 工具進行 KKday 的 SWOT 分析。本研究將使用情緒分析工具 VADER 來產生使用者評論的情緒分數，VADER 是 Python 程式套件，是一種基於詞典和規則的情感分析工具，並且特別適合用於社交媒體中表達的情感(Pandey, 2018)。

## 第二章 文獻探討

本章將目的地旅遊體驗平台，情緒分析，自然語言處理，以及 RE-SWOT 分析進行介紹及整理相關文獻。

### 第一節 目的地旅遊體驗平台(Local In-destination Tours and Activities Platform)

由於本研究以 Klook 為 KKday 的主要競爭者，針對 KKday 進行 SWOT 分析，因此本節將介紹 KKday 及 Klook 兩個亞洲的目的地旅遊體驗平台的歷程發展及相關研究文獻。

#### 一、KKday

KKday 成立於 2014 年，總部在台灣台北，為台灣成立的第一個專門提供目的地旅遊體驗與行程旅遊電商平台。KKday 提供旅人來自世界各地的多樣化在地體驗行程，讓旅行能更深入當地，在規劃上也更輕鬆、更自由。KKADY 目前平台服務遍布全球超過 80 個國家和 500 個城市、提供的旅遊行程與體驗超過 20,000 種，服務語言多元不只有繁簡中文，亦有英文、越語、泰文、日文、韓文等。全球 11 個市場皆設有辦公室與服務據點，包括台灣、香港、新加坡、日本、韓國、馬來西亞、泰國、菲律賓、越南、上海等地皆設立分公司。持續拓展北美、歐洲以及紐澳市場，並與亞洲萬里通結盟。平台致力於開發出不同地區不同類型的目的地旅遊體驗行程，提供方便簡易的預訂平台，並希望每個旅客，都能享受自由行的美好(KKday, 2019)。表 2-1 為 KKday 的大事紀。

表 2-1 KKday 的大事紀

年月	大事紀
2015.01	KKday 網站正式開站(姚舜，2015a)。
2015.03	推出第一屆最爽試玩員，目標為日韓地區(姚舜，2015b)。
2016.03	KKday 獲由 AppWorks Ventures 領投的 A 輪融資 450 萬美元(Maulani, 2016)。
2016.09	與亞洲萬里通合作(Frankie, 2016)。
2016.12	KKday 獲 A+ 輪融資 700 萬美元。(葉怡君，2016；INSIDE 硬塞的網路趨勢觀察, 2016)
2018.02	KKday 獲 B 輪策略融資 1050 萬美元，來自日本最大國際旅遊集團 H.I.S. 的資金挹注(Anny, 2018；吳元熙，2018)。
2018.06	與韓國超人氣卡通 POLI 波力合作，在台灣獨家推出的親子專車九份、野柳、十分生態之旅(莊士賢，2018)。
2018.11	KKday 獲 B+ 輪 LINE Ventures 和阿里巴巴台灣創業者基金策略融資(INSIDE 硬塞的網路趨勢觀察，2018；何佩珊，2018b；TechNews, 2018)。
2019.08	KKday 宣佈，獲得日本 Cool Japan Fund 策略投資(陳君毅，2019)。

## 二、Klook 客路

Klook 客路 2014 年成立於香港，其服務涵蓋世界各地共 25 國家和 250 多個城市，提供 8 萬多種景點、特色活動、交通及美食預訂選擇，累積超過 500 萬則真實旅客評論，並透過支援 41 種貨幣與 13 種語言，包含英文、越語、泰文、日文、韓文、簡中文、繁中文、印尼語、法文、西班牙文、德文、義大利文、俄文。Klook 已在香港、台北、新加坡、首爾、深圳、曼谷、吉隆坡和馬尼拉等 20 個城市設有辦公室，全球團隊有超過 1000 人，Klook 所推出的體驗行程標榜通過世界各地的旅遊體驗師親自體驗的行程，並且嚴格篩選當地合作的旅遊營運業者。表 2-2 為 Klook 的大事紀。

表 2-2 Klook 的大事紀

年月	大事紀
2014.09	Klook 客路網站上線。
2015.05	Klook 客路 APP 全球同步上線。上線一周，獲亞太 14 個國家及地區 App Store 推薦為「最佳新應用」(楊倩蓉，2016)。
2015.06	Klook 客路宣布完成由騰訊集團高級執行副總裁、前騰訊電商 CEO、天使投資人吳霄光領投的 150 萬美元的種子投資(極客公園，2015)。
2015.10	Klook 客路宣布獲得經緯創投(Matrix Partners)、華創資本及香港上市公司商會主席梁伯韜所領投的 A 輪融資共計 500 萬美元(張旭宏，2015)。

2016	Klook 客路網站及 App 在全球的行程預訂量也已超過 500 萬次。
2017.03	Klook 客路獲得由紅杉資本(Sequoia)領投的 3000 萬美元 B 輪投資。原 A 輪投資者經緯創投、微光資本(Welight)則繼續增投。宣布正式進入台灣市場(吳元熙，2017a；明報財經，2017)。
2017.10	Klook 客路宣布獲得 18 億新台幣 C 輪融資，由紅杉資本、高盛、經緯創投聯合領投，將把新一輪資金用在全球拓展，其中的 5 億資金，將全數投入台灣市場(吳元熙，2017b)。
2018.08	Klook 客路宣布獲 2 億美元的 D 輪融資，該輪融資由紅杉資本、TCV、經緯創投、高盛等，將加速拓展歐美市場(陳君毅，2018)。
2019.04	Klook 客路宣布獲 2.25 億美元的 D+輪融資，該輪融資由日本軟銀願景基金領投，紅杉資本(Sequoia)、經緯創投、TCV 及 OurCrowd 等跟投(Klook，2019)。

國內有部份學者針對 KKday 和 Klook 這兩個目的地旅遊體驗平台進行相關研究，呂維苓與黃仕斌(2016)透過實地觀察與訪問了解 KKday 的成功要因，讓更多想投入體驗經濟的企業作為參考；徐雅琪(2016)將 KKday 做為個案對象，分析平台的營運模式，再藉以訪談實際使用者的瀏覽、選購和活動體驗過程，欲建構出一個創新目的地旅遊活動平台的商業模式；侯姿伶(2018)以 KKday 的消費者為對象，進行問卷調查，驗證創新服務特性對於

功能價值及情感價值的影響；張文騰(2017)將 Klook 作為個案，利用價值網來探討平台的創新商業模式。上述研究皆是以 KKday 或 Klook 為單一研究對象，並且沒有競爭力分析的相關研究，因此本研究將使用 RE-SWOT 工具，分析 KKday 的優勢(strength)與劣勢(weakness)，並以 Klook 為競爭對手，分析 KKday 面臨的機會(opportunity)與威脅(threat)。

## 第二節 情緒分析(Sentiment Analysis)

### 一、情緒分析

情緒分析(Sentiment Analysis)也被稱為意見挖掘(Opinion Mining)，是指使用自然語言處理(NLP)、文本分析以及計算機語言學等技術，來系統地識別、提取、量化文本訊息中的主觀意見。情緒分析是利用計算處理方法來評估一個文本撰寫者或作者的主觀性態度、評價及情感(Pandey, 2018)。

情感分析是意見偵測的一種方法，意見的組成包含：意見的來源(亦即發表意見的人)、意見的目標個體、意見的型態(亦即極性(polarity))、含有意見的文字。最簡單的情感分析，僅單純判斷意見的型態，亦即正向、負向、還是中立；更進一步的情感分析，則會針對意見極性的強度給予量化的值，更進階的意見分析則會試圖判斷意見的來源和提及的目標個體，甚至會分析評論意見是提及目標個體的哪個面向，以及意見中隱含的情緒(Sentiment Analysis, n. d.)。舉例而言，假設有一則評論如下：

這台無線吸塵器電池續航力很不錯，充一次電可以用很久。

Posted by 李明明 2020/6/1

這則評論的意見來源為李明明，意見提及的目標個體為無線吸塵器，並且針對「電池續航力」這個面向表示意見，而所表示出的意見型態是正向的。

情緒分析相關研究應用範圍廣泛，涵蓋電影評論、餐飲業、旅遊觀光等文本。例如：蒐集 PTT 電影版的評論內容，透過情緒分析出形容詞的情緒分數，並且根據情緒分數推薦電影(張傳珩, 2017)。各項研究均從多樣網路文本中，取出特定議題的評論意見，以便分析其趨勢脈絡。目前也有一些現成的情緒分析工具，可以讀取文本訊息後，回傳訊息隱含的情緒極性(polarity)與強度，其中 VADER 是被廣泛使用的工具之一(Elbagir and Yang, 2019; Newman and Joyner, 2018; Hutto and Gilbert, 2014)。VADER 是一種基於詞典和規則的情感分析工具，特別適合用於社交媒體中表達的情感。VADER 被證實在處理社交媒體文本、《紐約時報》社論、電影評論和產品評論時，VADER 情緒分析的結果相較於其他情緒分析方法更為準確。VADER 不僅可以區分情緒極性(正面、負面或中性)，而且能給予情緒極性強度的量化值(Pandey, 2018)。

VADER 與傳統的情感分析方法相比的優勢，包括下列三點(Pandey, 2018)：

- (1) 在社交媒體類型的文本上效果非常好，且涵蓋多種不同的領域；
- (2) 不需要任何訓練數據，而是建構自一般化的、基於心理效價、人為制定的黃金標準情感詞典；
- (3) 效率夠快，可以用於線上的流動數據，並且不會嚴重影響速度與性能之間的平衡。

VADER 情緒分析結果會產生出四個數值，即為正向(positive)、負向(negative)、中性(neutral)和標準化加權綜合分數(compound)。標準化加權綜合分數(以下簡稱 Compound 值)是通過將情感詞典中每個單詞的情感分數總和，並根據規則進行調整，然後正規化為介於-1(極負向)和 1(極正向)之間的

分數(Swarnkar, 2020)。

## 二、旅遊情緒分析

情緒分析已廣為應用在旅遊產業，相關研究像是，Masrury et al. (2017)使用情感分析方法揭示對印度尼西亞最受歡迎的兩家 OTA (Online Travel Agent)平台 Traveloka 和 Tiket.com 中用戶評論的正面或負面情緒；Prameswari et al. (2017)通過情感分析獲得線上飯店評論句子中包含的情感信息；Kuhamanee et al. (2017)爬取線上社交網站的評論，來瞭解泰國曼谷的外國遊客，在泰國曼谷旅遊的體驗和情緒，以找出曼谷旅遊業可改善之處，Kuhamanee et al.使用了四種不同的方式做情緒分析，並且分析比較四種方式的準確率；Fu et al. (2019)則是以旅遊相關的新聞內容為分析對象，分別比較不同的情緒分析方法，以瞭解不同方法的預測準確性；Kirilenko et al. (2018)比較了旅遊中的自動情緒分析方法；Ramanathan and Meyyappan (2019)針對 Twitter 中阿曼王國旅遊反饋進行情緒分析。我們可以通過這些情緒分析應用於旅遊業的研究，看出旅遊業是近年情緒分析廣泛應用的領域之一。

## 第三節 自然語言處理(Nature Language Processing, NLP)

在本節中，我們將介紹在做分析前，準備使用數據應採取哪些 NLP 步驟來識別特徵。根據 Chowdhury (2003)的說法，自然語言處理(NLP)是一個研究和應用領域，它探索如何使用計算機來理解和影響自然語言的文本或言語來做有用的事情。Huang (2018)形容文本就像洋蔥需要一層一層剝開，文本則是要經過切成句子、斷成單詞、刪除停用詞的程序，最後才能用做文本分析。一般而言，文本訊息會經過下列預處理



的步驟(Parente, 2018)：

### 一、斷詞(Tokenization)：

Webster and Kit (1992)研究提及斷詞是 NLP 的起始步驟，斷詞有助於將文本訊息分為單個單詞。為了執行斷詞過程，Python 中有許多可用的開源工具，像是 NLTK Word Tokenize、Nlpdotnet Tokenizer 和 TextBlob Word Tokenize 等(Vijayarani and Janani, 2016)。

NLTK Word Tokenize 是著名的 Python Natural Language Processing Toolkit 工具包，NLTK 是 Python 中建構來用於研究人類語言相關數據的開源工具，它提供了超過 50 個語料庫和詞彙資源，例如 WordNet，以及用於分類、斷詞、詞幹提取、單詞標記、解析和語義分析的文本處理套件。

Nlpdotnet 是一個以神經網絡為基礎的 Python 庫，用於處理自然語言。它可以執行斷詞，詞性標記，語義標記及依賴項解析。儘管斷詞看似只是其中一項，但其為相當重要的步驟，幾乎所有自然語言處理都會需要用到。

TextBlob 是用於處理文本數據的 Python 庫，用於研究常見的自然語言處理(NLP)，像是文本挖掘、文本分析和文本處理。它提供了 API 可以連結使用各種工具包，例如斷詞、詞性標記、名詞短語提取、情感分析、分類、翻譯等。

### 二、大寫轉小寫(Lowercase transformation)

將所有單詞都統一轉換為小寫形式，以避免單詞因大小寫的不同，而被辨識為不同的單詞。例如，“KKDAY”、“KKday”、“kkday”三個單詞其實都是一樣的，可將其全部轉成小寫的“kkday”。

### 三、停用詞刪除(Stopword removal)

由於一些常用字或詞使用的頻率相當的高，英語中比如

a、the、he 等，每個文本幾乎都包含了這些字詞，如果在進行分析時將這些詞一起計算，那麼這些詞會因為出現頻率較高有較高的機率被判定為關鍵字，所以一般把這些停用詞事先刪除，再在分析時不使用。

一般有兩種常用的停用詞刪除方法：

#### (一)經典方法：

經典方法是以停用詞語料庫中獲得的停用詞為基礎做停用詞刪除。

#### (二)Z 方法：

除了經典方法的停用詞語料庫，還有使用 Zipf's 定律的停用詞刪除方法，刪除包括：最頻繁的詞(TF-High)、只一次出現的單詞，即單例單詞(TF1)以及逆向文件頻率(IDF)較低的單詞。

### 四、詞性標記(part-of-speech tagging, POS tagging)

每個單詞有不同的詞性，例如：名詞，動詞，冠詞，形容詞，介詞，代詞，副詞，連詞和感嘆詞，確定單詞的詞性能更準確了解文本。執行自然語言處理時，會需要將單詞進行分類以及將對應的單詞進行標記，這過程稱為詞性標記，詞性標註的正確與否將會直接影響到後續的字詞、句法和語義等分析。它是自然語言處理中重要的基礎研究之一，已被廣泛的應用於機器翻譯、文字識別、語音識別和信息檢索等領域。Tsuruoka et al. (2015)為了生物醫學文本，開發了健全的詞性標註工具。

### 五、詞性還原(Lemmatization)

詞形還原(Lemmatization)是文本預處理中的重要部分，詞形還原就是去掉單詞的詞綴，提取單詞的主幹部分，通常

提取後的單詞會是字典中的單詞，不同於詞幹提取(stemming)，詞幹提取後的單詞則不一定會出現在字典中。例如，將“change”、“changing”、“changes”、“changed”和“changer”，都將其詞形還原為“change”；若做詞幹提取則是“chang”。

## 六、詞組搭配(Collocations)

詞組搭配指的是兩個語言單位的組合，如兩個單詞  $w_1$  和  $w_2$  同時出現，通常採用統計數據來進行識別，識別  $w_1$  和  $w_2$  是否是較常同時出現的單詞詞組。典型的統計方法有： $t$  分數(t-score, TSC)、皮爾森卡方檢定(Pearson's  $\chi$ -square test,  $\chi^2$ )、似然比(likelihood ratio, LR)和點式交互資訊(pointwise mutual information, PMI)。Thanopoulos et al. (2002)比較了幾種詞組搭配識別方法，根據其研究的實驗結果，概似比(LR)的方法效果是相對較好的。似然比的計算方式如下：

$$-2 \log \lambda = 2 \cdot \log \frac{L(H_1)}{L(H_0)} \quad (2-1)$$

其中  $H_0$  和  $H_1$  為成對的假設， $H_0$  是假設  $w_2$  的出現與  $w_1$  先前的出現兩者是獨立事件， $H_1$  是假設  $w_2$  的出現與  $w_1$  先前的出現兩者是相依事件， $L(H)$  代表假設  $H$  基於觀測資料推得的可能性，單詞的機率是假設服從二項分佈(Dunning, 1993)。

## 第四節 RE-SWOT 分析(Requirements Engineering SWOT)

SWOT 分析是 80 年代初，由美國舊金山大學的管理學教授韋里克提出，用於分析企業的優勢(Strengths)、劣勢(Weaknesses)、機會(Opportunities)和威脅(Threats)，如表 2-3，分析後的結果常被做為企業戰略制定的參考在 SWOT 中，優勢是期望績效的增強因

素，而劣勢是期望績效的抑制因素，兩者都是組織可控內部的因素，用以考量企業內部條件的優勢和劣勢，是否有利於在產業內競爭。另外，機會與威脅則是組織無法控制的外部環境增強因素及抑制因素，這些外部環境因素是針對企業外部環境進行探索，瞭解產業未來情勢之演變。確定了這四個因素後，利益相關者將對其有戰略的瞭解，並制訂行動計劃以改善自身的劣勢、利用機會、保持優勢並防止威脅。

表 2-3 SWOT 分析矩陣

	內部因素	外部因素
有利的	優勢 Strengths	機會 Opportunities
有害的	劣勢 Weaknesses	威脅 Threats

Parente (2018) 提出一種以 SWOT 分析進行需求工程 (Requirements Engineering) 的方法，所謂需求工程是指應用有效的技術進行需求分析，以幫助企業理解確定客戶的需求。Parente 認為用戶評論是瞭解客戶需求的重要依據，因此設計了一套流程(如圖 2-1)，從產品的用戶評論中，擷取用戶提及的產品特徵及對這些特徵的評價，以自動化生成 SWOT 矩陣，並據以進行需求工程，因此 Parente 將此流程建置的 SWOT 矩陣稱為 RE-SWOT (Requirements Engineering SWOT)。在 Parente 的研究中，是以手機 app 為需求工程的研究對象，由於本研究會應用到 Parente 的 RE-SWOT 工具，因此以下針對圖 2-1 的流程進行詳細介紹。

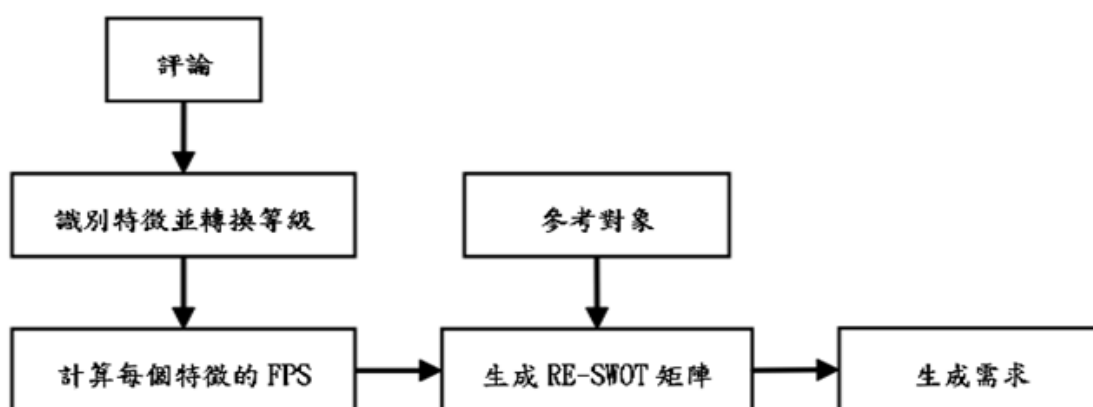


圖 2-1 RE-SWOT 流程圖(Parente, 2018)

### 一、識別特徵並轉換等級

爬取用戶在 App 商店中對於手機 App 的留言評論及評分，所有的評論會經過第三節所提到的預處理步驟，以識別出被用戶提及的應用程序的特徵；此外，使用者評分的範圍為 1 到 5，會被平移轉換為-2 到+2 的範圍。這種轉換的目的是在後續計算特徵分數(Feature Performance Score, FPS)時，為評論指定正面或負面的方向。

### 二、計算每個特徵的 FPS

給定一組應用程序  $A = \{a_1, a_2, a_i, \dots, a_m\}$  和一組特徵詞  $F = \{f_1, f_2, f_j, \dots, f_n\}$ ，則可以用公式 2-2 計算與特徵詞  $j$  相關的應用程序  $i$  的 FPS。

$$FPS_{i,j} = \frac{S_{i,j} \cdot V_{i,j}}{\sum_{i=1}^m |S_{i,j} V_{i,j}|} \quad (2-2)$$

其中  $S_{i,j}$  代表應用程序  $i$  中特徵詞  $j$  的使用者情感評分，其計算方式為提及該特徵詞的評論的使用者評分之總和，除以使用者評分最大可能總和，例如，如果在兩個 5 星級評論和兩個 2 星級評論中提到某個特徵，則該特徵的特徵分數等於

$(2 + 2 - 1 - 1)/(2 + 2 + 2 + 2) = +0.25$ ;  $V_{ij}$  代表應用程式  $i$  中提及特徵詞  $j$  的用戶評論數量，例如，如果某個應用程式收到 2 條評論「上傳照片時應用程式癱瘓。每次嘗試上傳照片時，都會出現錯誤消息。」和「應用程式最近癱瘓了」，其中只有一條評論提及「上傳照片」這個特徵，因此「上傳照片」的特徵量為 1 (Parente, 2018)。

### 三、生成 RE-SWOT 矩陣

對於每個特徵詞，將根據三個標準對每個應用程式的分數進行評估：第一，是否為正向或負向；第二，分數高於或低於該特徵的平均分數，第三，是否為競爭對手的應用程式特徵。第一個標準，由下列兩公式判定。

$$FPS_{i,j} \geq \sigma \quad (2-3)$$

$$FPS_{i,j} \leq -\sigma \quad (2-4)$$

如果公式 2-3 成立，則認為 FPS 為正；如果公式 2-4 成立，則認為 FPS 為負。因此，當 FPS 在  $[-\sigma, +\sigma]$  範圍內時，它被稱為中性。

第二個標準，FPS 用於確定某項特徵是否是獨特的，或者是否高於或低於市場平均水平。如果公式 2-5 成立，則 FPS 高於市場平均水平；如果公式 2-6 成立，則 FPS 低於市場平均水平。

$$FPS_{i,j} - \overline{FPS}_l > \sigma \quad (2-5)$$

$$FPS_{i,j} - \overline{FPS}_l < -\sigma \quad (2-6)$$

FPS 的值為 +1 或 -1 時，代表該特徵是獨特的，即市場上沒有其他應用有此特徵。對第三步中的公式，Parente 的研究採用  $\sigma = 0.1$ 。

第三個標準，評估特徵詞是否屬於競爭對手及比較市場

水平。如為競爭對手且表現優於市場平均水平的特徵被歸類為威脅；反之，若分數為負向且低於平均值的功能，則表示機會。如果是指自己的應用程序中的特徵，則為優勢(正面高於平均值時)或劣勢(負面低於平均值時)。特徵若不符合上述三個標準則不會出現在 RE-SWOT 矩陣中。

最後一步，將確定優勢、劣勢、威脅和機會的列表，產生需求以解決這些問題。在這一步中，將應用 TOWS 框架(Wehrich, 1982)，該框架結合了 SWOT 分析的結果如表 2-4，研究對象的特徵詞將被填入自身特徵欄，分別針對其特徵詞所歸類出的結果生成 RE-SWOT 矩陣，RE-SWOT 矩陣可以判斷出其需求，如果是優勢則應保留或擴展的特徵，劣勢則應最小化問題和錯誤特徵；競爭對手的特徵詞將被填入競爭對手特徵欄，如為威脅則是可被效仿的特徵，機會則是可以被開發的新特徵。此表可以確定提高公司競爭力的策略，得出在保持或提高此類應用競爭力的新需求(Parente, 2018)。

表 2-4 RE-SWOT 分析矩陣(Parente,2018)

	自身特徵	競爭對手特徵
正面的	S:應保留或擴展的特徵	T:可被效仿的特徵
負面的	W:應最小化問題和錯誤特徵	O:可以被開發的新特徵

舉例而言，假設以 App1 為研究對象，並以 App2 和 App3 為其競爭對手，根據圖 2-1 的流程，生成如表 2-4 的 RE-SWOT 分析矩陣。

表 2-5 RE-SWOT 分析矩陣範例(Parente, 2018)

		特徵詞	
		App 1	App 2、App 3
性能	正面的和高於市場平均水平	產品訂購	介面設計(App 2) 系統同步(App 2) 瀏覽紀錄(App 3)
	負面的和低於市場平均水平	瀏覽紀錄	產品訂購(App 2) 瀏覽紀錄(App 2) 客服(App 3)

從表 2-5 結果可以看出，App 1 的特徵產品訂購的 FPS 為正面的且高於市場平均水平；App 1 的特徵瀏覽紀錄的 FPS 為負面的且低於市場平均水平；競爭對手 App 2 及 App 3 的特徵介面設計、系統同步以及瀏覽紀錄的 FPS 為正面的且高於市場平均水平；App 2 及 App 3 的特徵產品訂購、瀏覽紀錄以及客服的 FPS 為負面的且低於市場平均水平。

根據表 2-5，App1 便可以制定四個策略：

#### (一)WT 策略：

指在最小化劣勢以使其不易受到威脅影響的策略要求，例如降低瀏覽紀錄的功能對使用者的不便，以防使用者轉向使用 App3。

#### (二)WO 策略：

指在克服劣勢以尋求機會的策略要求，例如提升瀏覽紀錄的功能，使其比 App2 吸引使用者使用。



### **(三)ST 策略：**

指在利用優勢降低威脅脆弱性的策略要求，例如在利用產品訂購簡單方便，來弱化使用者對介面設計的要求等。

### **(四)SO 策略：**

指在尋求契合優勢的機會的策略要求，例如在強化發展產品訂購功能，可以吸引轉移 App2 對有產品訂購要求的使用者。

本研究將以 Parente (2018)的方法去分析 KKday 及 Klook 的競爭力關係，但在建立 RE-SWOT 需要使用者評分，只有目的地旅遊體驗平台官網上有使用者評分，而 Tripadvisor 評論中並沒有使用者評分，本研究將以情緒分析產生的情緒分數，做為生成 RE-SWOT 矩陣需要的使用者評分。



## 第三章 研究方法

本章將介紹研究的流程架構，包含資料蒐集、資料預處理、情緒分析以及競爭力分析。

### 第一節 研究架構

為了解目的地旅遊體驗平台的競爭力分析，來提供 KKday 平台的發展方向，本研究將使用 Tripadvisor 網站上 KKday 和 Klook 相關產品的線上評論，並用競爭力分析找出提升目的地旅遊體驗平台 KKday 產品競爭力的需求策略，如圖 3-1 所示，它總結了本研究流程。

本文首先使用網路爬蟲從 Tripadvisor 網站中提取線上評論。然後，參考 Parente (2018) 研究進行競爭力分析，首先對數據進行預處理的方法，使用自然語言處理(NLP)和 Python 程式語言來構建評論數據集，包括數據清理、斷詞、大寫轉小寫、停用詞刪除、詞性標記和詞性還原。

另外，等級部分先將 Tripadvisor 評論使用 VADER 情緒分析產生情緒分數後，利用 Elbagir and Yang (2019) 研究中，等級轉換方法，將情緒分數轉換為 $[-2,2]$ 之間的等級。最後，結合兩者，計算 FPS 分數並產生 RE-SWOT 矩陣，生成 KKday 的競爭需求。

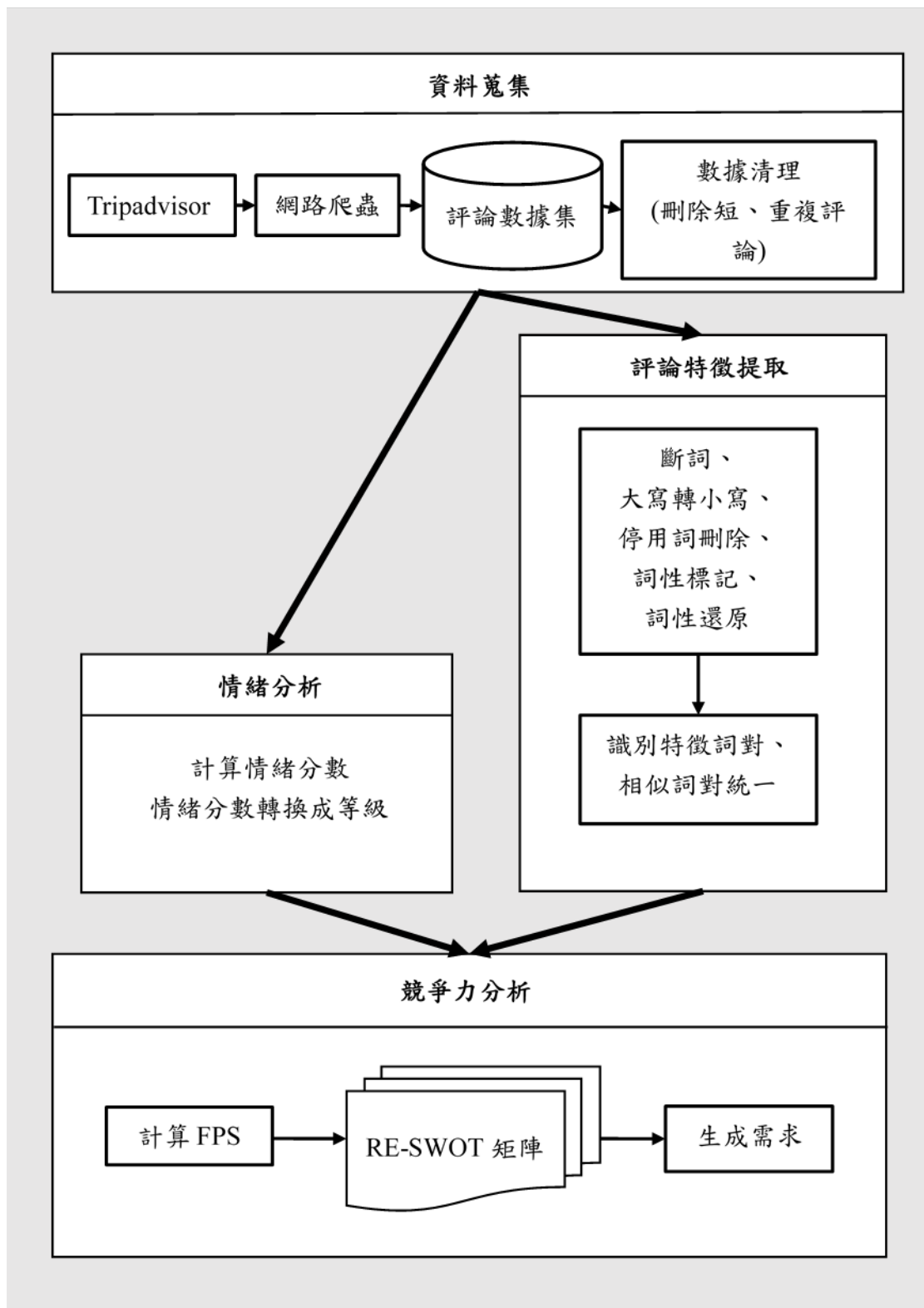


圖 3-1 研究流程圖

## 一、資料蒐集(Data collection)

現今有許多研究，會利用網路爬蟲的方式從網路上的平台蒐集文本數據(游綉雯, 2015; 郭炳良, 2017; 李佳孟, 2019)，本研究的數據，將使用 Python 程式語言撰寫網路爬蟲功能，從 Tripadvisor 上爬取針對 KKday 及 Klook 的英文使用者評論。

由於資料蒐集以及轉移過程中，可能造成資料格式錯誤或導致資料遺失等問題發生，因此在做資料分析資料前先透過數據清理的方式將資料前置處理，以方便正確地做資料分析(Huang, 2016)。

數據清理亦可檢測和刪除線上文本數據中的不準確或無用的記錄，例如拼寫錯誤和非目標語言，僅留下有價值的旅遊相關訊息。在本研究中，我們預計先刪除了文字長度較短的評論，再刪除由單個用戶發布的多次重複評論，因為重複評論可能會導致統計偏差，因此我們在用戶的重複記錄中僅保留了一條記錄。以確保數據樣本的真實性和準確性(Hou et al., 2019)。

以 Tripadvisor 上爬取 KKday 中的產品評論其中三則為例「Is it available as the date we booking from kkday?」、「I buy ticket in KKdays, and they will sent me QR code through email. Use QR code to go in, very convenience, easy, and save time.」、「Nice day out to Shifen with KKday tours and it was cool going to the Shifen Waterfall on a hot day. Recommended.」，其中「Is it available as the date we booking from kkday?」為長度較短的文句且非評論的無用記錄，因此將會被刪除以確保文本數據的準確性。

## 二、評論特徵提取

參考 Parente (2018)研究對數據進行預處理的方法，使用

NLTK 和 Python 編程語言來做評論特徵提取，包括數據清理、斷詞、大寫轉小寫、停用詞刪除、詞性標記和詞性還原。本研究將使用 NLTK(Natural Language Tool Kit)，他是一套基於 Python 的自然語言處理工具。

### **(一)斷詞(Tokenization)：**

斷詞是將文本串序列分解為單詞、關鍵字、短語、符號和其他稱為詞的元素之類的動作。斷詞可以是單個單詞、短語甚至整個句子。在斷詞過程中，某些字符(如標點符號)被丟棄。以其中一則評論為例，「I buy ticket in KKdays, and they will sent me QR code through email. Use QR code to go in, very convenience, easy, and save time.」，可以被斷句成，「I buy ticket in KKdays, and they will sent me QR code through email.」、「Use QR code to go in, very convenience, easy, and save time.」，再將兩句進一步斷詞成，「I」、「buy」、「ticket」、「in」、「KKdays」、「,」、「and」、「they」、「will」、「sent」、「me」、「QR」、「code」、「through」、「email」、「.」、「Use」、「QR」、「code」、「to」、「go」、「in」、「,」、「very」、「convenience」、「,」、「easy」、「,」、「and」、「save」、「time」、「。」(Huang, 2018)。

### **(二)大寫轉小寫(Lowercase transformation)**

大寫轉小寫的作用為將所有斷詞以小寫形式以進行統一，可減少單詞因大小寫的不同而被辨識為不同的單詞。本研究將使用 Python 中的 `string.lower()` 可以將所有斷詞都轉換為小寫，以便統一以小寫的形式呈現。前一步驟中斷詞後的「I」、「KKdays」、「QR」有以大寫形式呈現的字母，其將轉換為小寫的「i」、「kkdays」、「qr」。

### **(三)停用詞刪除(Stopword removal)**

由於一些常用字或者詞使用的頻率相當的高，英語中比如 a, the, he 等，每個文本幾乎都包含了這些字詞，如果做分析時將它們當關鍵字進行計算，那麼所有這些都會被計算進去，而且沒有區分度，所以一般把這些詞直

接去掉，不可當做關鍵字。

在英語單詞中停用詞對句子沒有多大意義。在不改變句子含義的情況下，可以將其刪除並忽略它們。例如，a, the, and, but, 等的單詞，可以在停用詞的語料庫中獲取這些單詞。本研究將使用 Python NLTK 套件的英文版本之停用詞方法 stopwords.words('english')，其中英文版本的 stopwords 以代名詞、介系詞為主，總共 179 個停用字。

將轉換完成的的詞「i」、「buy」、「ticket」、「in」、「kkdays」、「,」、「and」、「they」、「will」、「sent」、「me」、「qr」、「code」、「through」、「email」、「.」、「Use」、「qr」、「code」、「to」、「go」、「in」、「,」、「very」、「convenience」、「,」、「easy」、「,」、「and」、「save」、「time」、「.」，進行停用詞刪除後結果為「buy」、「ticket」、「kkdays」、「sent」、「qr」、「code」、「email」、「qr」、「code」、「go」、「convenience」、「easy」、「save」、「time」。

#### (四)詞性標記(part-of-speech tagging, POS tagging)

特徵通常以名詞，動詞和形容詞來描述，因此，我們使用 Python NLTK 中的 pos\_tag 方法(part-of-speech tagging)來對單詞的詞性進行標記，標記後的結果是二元陣列格式。之後從這個二元數列中挑出我們所有需要的詞性(名詞、動詞、形容詞)，對應詞性如表 3-1 所示(*Part of Speech Tagging with NLTK*, n. d. )。再存放進一個二元數列。

表 3-1 詞性列表

詞性標記	說明	範例
NN	名詞單數形式	desk
NNS	名詞複數形式	desks
NNP	專有名詞	Harrison
NNPS	專有名詞複數形式	Americans
VB	動詞	take

VBD	動詞，過去時；過去式	took
VBG	動詞，動名詞/現在分詞	taking
VDN	動詞，過去分詞	taken
VBP	動詞，現在	take
VBZ	動詞，第三人稱	takes
JJ	形容詞	big
JJR	(形容詞或副詞的) 比較級形式	bigger
JJS	(形容詞或副詞的) 最高級	biggest

將完成停用詞刪除後的詞「buy」、「ticket」、「kkdays」、「sent」、「qr」、「code」、「email」、「qr」、「code」、「go」、「convenience」、「easy」、「save」、「time」做詞性標記。標記後的結果是二元陣列格式「buy, VB」、「ticket, NN」、「kkdays, NNS」、「sent, VB」、「qr, NN」、「code, NN」、「email, NN」、「qr, NN」、「code, NN」、「go, VB」、「convenience, NN」、「easy, JJ」、「save, VB」、「time, NN」。

### (五)詞性還原(Lemmatization)

本研究利用 Python NLTK 中的 Lemmatizer()方法將標記好的單詞還原成原型。此步驟是為了減少單詞詞性變化造成的影響，否則可能產生出多餘的相同特徵。

詞性標記結果其中之一的單詞「kkdays」將做詞性還原成「kkday, NN」

### (六)詞組搭配(Collocations)

詞組是通常同時出現多個單詞組合而成的表達方式。例如，單詞「door」、「chair」，這兩單詞比較不會同時在文本中出現；單詞「open」、「door」，這兩單詞比較長會同時在文本中出現，因此會將「open」、「door」作為一個詞組。

本研究將參考 Thanopoulos et al.(2002)研究結果，他比較了幾種詞組搭配(Collocations)方法，根據其研究的實驗結果，利用 Likelihood ratio 的方法效果是最好的，所

以本研究將採用 Likelihood ratio 的計算方式進行。並且將所有評論中出現三次以上的詞組，選前 20 詞對作為特徵詞對。

將詞性還原的步驟結果「buy, VB」、「ticket, NN」、「kkday, NN」、「sent, VB」、「qr, NN」、「code, NN」、「email, NN」、「qr, NN」、「code, NN」、「go, VB」、「convenience, NN」、「easy, JJ」、「save, VB」、「time, NN」利用 Likelihood ratio 的方法做詞組的搭配可以搭配出(qr, code)。

### (七)相似詞對統一(merge similar features)

本研究利用人工方式將相似有關聯的詞對。統一使用為相同的詞對。例如詞組「high speed」和「speed rail」都是描述高鐵，所以我們將其統一使用「high speed」。

## 三、情緒分析

本研究使用情緒分析工具 VADER (Valence Aware Dictionary and sEntiment Reasoner)，針對從 Tripadvisor 爬取的評論進行分析，VADER 分析後會輸出四項資訊：Compound 值、positive 的可能性、negative 的可能性、中立的可能性，本研究會引用 Elbagir and Yang (2019)提出的方法，將 VADER 輸出的資訊轉換為[-2,2]之間的等級分數。

### (一) 情緒分數

本研究首先利用 VADER 情緒分析產生的 Compound 值，依 Elbagir and Yang (2019)所提出的方法中，將 Compound 值依不同的門檻值(threshold)，將評論依下列規則分類為正向(1)，負向(-1)或中性(0)：

正向：Compound 值 > 0.001，情緒分數 = 1

中性：0.001 > Compound 值 > -0.001，情緒分數 = 0

負向：Compound 值 < -0.001，情緒分數 = -1

### (二) 情緒分數轉換成等級

本研究根據 Elbagir and Yang (2019)研究中，使用的



等級分數轉換方法，將情緒分數轉換為[-2,2]之間的等級。Elbagir and Yang 定義了一個評分規則，進一步將情感極性的強度分成五類：極正向(2)，正向(1)，中性(0)，負向(-1)和極負向(-2)，評分規則設置如下：

If (情緒分數) = 1:

If (positive 可能性 > 0.5): 等級 = +2

Else: (positive 可能性 < 0.5): 等級 = +1

If (情緒分數) = -1:

If (negative 可能性 > 0.5): 等級 = -2

Else: (negative 可能性 < 0.5): 等級 = -1

If (情緒分數 = 0): 等級 = 0

例如：若有一則評論「I buy ticket in KKdays, and they will sent me QR code through email. Use QR code to go in, very convenience, easy, and save time.」，經過 VDAER 分析後的結果得出四個 negative = 0.000、neutral = 0.783、positive = 0.217、Compound = 0.7490。由於 Compound 值  $0.7490 > 0.001$ ，情緒分數 = 1，再依據評分規則，情緒分數 = 1 時，比較 positive 來轉換等級，因為 positive 0.217 < 0.5，所以等級為+1。

#### 四、RE-SWOT 分析

依據第二章描述的 FPS 計算方式，計算各特徵的 FPS。舉例而言，假設特徵 qr code 在 kkday 出現六次，等級分別為 1、1、-1、-1、1、1，qr code 在 klook 出現四次，等級分別為 1、1、-1、-1。通過公式 2-2 的計算結果 qr code 在 kkday 的 FPS 為 1.0，在 klook 的 FPS 為 0。假設  $\sigma = 0.1$ ，根據公式 2-3 判定 qr code 為正向的。且 qr code 在 kkday 的 FPS 值為+1，代表該特徵是獨特的，即市場上沒有其他應用有此特徵。因為是 kkday 自身特徵，所以 qr code 會被放在 RE-SWOT 分析矩陣的優勢欄中，如表 3-2 所示。

表 3-2 RE-SWOT 分析矩陣範例

		特徵詞	
		kkday	klook
性能	正面的和高於 市場平均水平	S 優勢 qr code	T 威脅
	負面的和低於 市場平均水平	W 劣勢	O 機會



## 第四章 研究結果

### 第一節 數據來源

首先，本研究爬取 TripAdvisor 中與 KKday 及 Klook 相關的產品及服務的評論，並將收集到的評論，通過第三章所述之方法進行計算處理，先將重複及過短的評論刪除，刪除後 KKday 的評論數量共 206 則，Klook 共 158 則，再進行斷詞、大寫轉小寫、刪除停用詞、詞性標記及詞性還原工作。

### 第二節 特徵詞組

從 KKday 及 Klook 產生出的 20 個詞組，如表 4-1 所示。通過刪除兩邊重複的詞對，以及相似詞對統一的方式，最後將 20 個詞對選作特徵詞組，如表 4-2。



表 4-1 KKday 及 Klook 評論產生詞組

	<b>KKday</b>	<b>Klook</b>
1	('tour', 'guide')	('front', 'desk')
2	('moon', 'lake')	('observation', 'deck')
3	('observation', 'deck')	('hot', 'spring')
4	('sun', 'moon')	('sky', 'lantern')
5	('pocket', 'wifi')	('mrt', 'station')
6	('credit', 'card')	('night', 'market')
7	('theme', 'park')	('rock', 'formation')
8	('qr', 'code')	('main', 'station')
9	('leofoo', 'village')	('high', 'speed')
10	('day', 'tour')	('taipei', 'main')
11	('guo', 'kuang')	('theme', 'park')
12	('sim', 'card')	('food', 'court')
13	('high', 'speed')	('sim', 'card')
14	('bus', 'stop')	('life', 'saver')
15	('food', 'court')	('ximen', 'station')
16	('souvenir', 'shop')	('speed', 'rail')
17	('shuttle', 'bus')	('red', 'line')
18	('pineapple', 'cake')	('shuttle', 'bus')
19	('rail', 'bike')	('qr', 'code')
20	('cable', 'car')	('priority', 'pass')

表 4-2 選用之特徵詞組

	特徵詞組
1	tour guide
2	sun moon lake
3	observation deck
4	pocket wifi
5	credit card
6	theme park
7	qr code
8	day tour
9	bus
10	sim card
11	high speed
12	food court
13	souvenir shop
14	rail bike
15	cable car
16	front desk
17	hot spring
18	sky lantern
19	mrt station
20	rock formation

### 第三節 特徵 FPS

計算 FPS 時，需要輸入使用者評分，但在 Tripadvisor 爬取的使用者評論，並未附帶使用者的評分，因此本研究使用 VADER 情緒分析工具對評論進行情緒分析，利用產生的情緒分數做為使用者評分的情感等級。RE-SWOT 矩陣的結果會與 $\sigma$ 值的設定有關，本研究共實驗了四種 $\sigma$ 值：0.1、0.2、0.3、0.4，當 $\sigma = 0.1$ 時，結果如表 4-3 所示；當 $\sigma = 0.2$ 時，結果如表 4-4 所示；當 $\sigma = 0.3$ 時，結果如表 4-5 所示；當 $\sigma = 0.4$ 時，結果如表 4-6 所示。

表 4-3 KKday 及 Klook 特徵之 FPS ( $\sigma = 0.1$ )

$\sigma = 0.1$		Reference	Competitor
		KKday(FPS)	Klook(FPS)
tour guide	FPS	0.82	0.18
	Situation	Pos, above average	Pos, below average
sun moon lake	FPS	-1.00	-
	Situation	Neg, unique	-
observation deck	FPS	0.67	0.33
	Situation	Pos, above average	Pos, below average
pocket wifi	FPS	0.82	0.18
	Situation	Pos, above average	Pos, below average
credit card	FPS	0.73	0.27
	Situation	Pos, above average	Pos, below average
theme park	FPS	0.60	0.40
	Situation	Pos, below average	Pos, below average
qr code	FPS	1.00	-
	Situation	Pos, unique	-

<b>day tour</b>	FPS	0.73	0.27
	Situation	Pos, above average	Pos, below average
<b>bus</b>	FPS	0.54	0.46
	Situation	Pos, below average	Pos
<b>sim card</b>	FPS	0.60	0.40
	Situation	Pos, below average	Pos, below average
<b>high speed</b>	FPS	0.29	0.71
	Situation	Pos, below average	Pos, above average
<b>food court</b>	FPS	0.27	0.73
	Situation	Pos, below average	Pos, above average
<b>souvenir shop</b>	FPS	0.62	0.38
	Situation	Pos, above average	Pos, below average
<b>rail bike</b>	FPS	1.00	-
	Situation	Pos, unique	-
<b>cable car</b>	FPS	1.00	-
	Situation	Pos, unique	-
<b>front desk</b>	FPS	0.29	0.71
	Situation	Pos, below average	Pos, above average
<b>hot spring</b>	FPS	-	1.00
	Situation	-	Pos, unique
<b>sky lantern</b>	FPS	0.43	0.57
	Situation	Pos, below average	Pos
<b>mrt station</b>	FPS	0.08	0.92
	Situation	Neu, below average	Pos, above average
<b>rock formation</b>	FPS	0.17	0.83
	Situation	Pos, below average	Pos, above average

<b>FPS AVG.</b>	0.51	0.52
-----------------	------	------

表 4-4 KKday 及 Klook 特徵之 FPS ( $\sigma = 0.2$ )

$\sigma = 0.2$		Reference	Competitor
		KKday(FPS)	Klook(FPS)
<b>tour guide</b>	FPS	0.82	0.18
	Situation	Pos, above average	Pos, below average
<b>sun moon lake</b>	FPS	-1.00	-
	Situation	Neg, unique	-
<b>observation deck</b>	FPS	0.67	0.33
	Situation	Pos, below average	Pos, below average
<b>pocket wifi</b>	FPS	0.82	0.18
	Situation	Pos, above average	Pos, below average
<b>credit card</b>	FPS	0.73	0.27
	Situation	Pos, above average	Pos, below average
<b>theme park</b>	FPS	0.60	0.40
	Situation	Pos, below average	Pos, below average
<b>qr code</b>	FPS	1.00	-
	Situation	Pos, unique	-
<b>day tour</b>	FPS	0.73	0.27
	Situation	Pos, above average	Pos, below average
<b>bus</b>	FPS	0.54	0.46
	Situation	Pos, below average	Pos
<b>sim card</b>	FPS	0.60	0.40
	Situation	Pos, below average	Pos, below average
<b>high speed</b>	FPS	0.29	0.71



	Situation	Pos, below average	Pos, below average
<b>food court</b>	FPS	0.27	0.73
	Situation	Pos, below average	Pos, above average
<b>souvenir shop</b>	FPS	0.62	0.38
	Situation	Pos, below average	Pos, below average
<b>rail bike</b>	FPS	1.00	-
	Situation	Pos, unique	-
<b>cable car</b>	FPS	1.00	-
	Situation	Pos, unique	-
<b>front desk</b>	FPS	0.29	0.71
	Situation	Pos, below average	Pos, below average
<b>hot spring</b>	FPS	-	1.00
	Situation	-	Pos, unique
<b>sky lantern</b>	FPS	0.43	0.57
	Situation	Pos, below average	Pos
<b>mrt station</b>	FPS	0.08	0.92
	Situation	Neu, below average	Pos, above average
<b>rock formation</b>	FPS	0.17	0.83
	Situation	Pos, below average	Pos, above average
<b>FPS AVG.</b>		0.51	0.52

表 4-5 KKday 及 Klook 特徵之 FPS ( $\sigma = 0.3$ )

$\sigma = 0.3$		Reference	Competitor
		KKday(FPS)	Klook(FPS)
<b>tour guide</b>	FPS	0.82	0.18
	Situation	Pos, above average	Pos, below average

<b>sun moon lake</b>	FPS	-1.00	-
	Situation	Neg, unique	-
<b>observation deck</b>	FPS	0.67	0.33
	Situation	Pos, below average	Pos, below average
<b>pocket wifi</b>	FPS	0.82	0.18
	Situation	Pos, above average	Pos, below average
<b>credit card</b>	FPS	0.73	0.27
	Situation	Pos, below average	Pos, below average
<b>theme park</b>	FPS	0.60	0.40
	Situation	Pos, below average	Pos, below average
<b>qr code</b>	FPS	1.00	-
	Situation	Pos, unique	-
<b>day tour</b>	FPS	0.73	0.27
	Situation	Pos, below average	Pos, below average
<b>bus</b>	FPS	0.54	0.46
	Situation	Pos, below average	Pos
<b>sim card</b>	FPS	0.60	0.40
	Situation	Pos, below average	Pos, below average
<b>high speed</b>	FPS	0.29	0.71
	Situation	Pos, below average	Pos, below average
<b>food court</b>	FPS	0.27	0.73
	Situation	Pos, below average	Pos, below average
<b>souvenir shop</b>	FPS	0.62	0.38
	Situation	Pos, below average	Pos, below average
<b>rail bike</b>	FPS	1.00	-
	Situation	Pos, unique	-

<b>cable car</b>	FPS	1.00	-
	Situation	Pos, unique	-
<b>front desk</b>	FPS	0.29	0.71
	Situation	Pos, below average	Pos, below average
<b>hot spring</b>	FPS	-	1.00
	Situation	-	Pos, unique
<b>sky lantern</b>	FPS	0.43	0.57
	Situation	Pos, below average	Pos
<b>mrt station</b>	FPS	0.08	0.92
	Situation	Neu, below average	Pos, above average
<b>rock formation</b>	FPS	0.17	0.83
	Situation	Pos, below average	Pos, above average
<b>FPS AVG.</b>		0.51	0.52

表 4-6 KKday 及 Klook 特徵之 FPS ( $\sigma = 0.4$ )

$\sigma = 0.4$		Reference	Competitor
		KKday(FPS)	Klook(FPS)
<b>tour guide</b>	FPS	0.82	0.18
	Situation	Pos, below average	Pos, below average
<b>sun moon lake</b>	FPS	-1.00	-
	Situation	Neg, unique	-
<b>observation deck</b>	FPS	0.67	0.33
	Situation	Pos, below average	Pos, below average
<b>pocket wifi</b>	FPS	0.82	0.18
	Situation	Pos, below average	Pos, below average
<b>credit card</b>	FPS	0.73	0.27

	Situation	Pos, below average	Pos, below average
<b>theme park</b>	FPS	0.60	0.40
	Situation	Pos, below average	Pos, below average
<b>qr code</b>	FPS	1.00	-
	Situation	Pos, unique	-
<b>day tour</b>	FPS	0.73	0.27
	Situation	Pos, below average	Pos, below average
<b>bus</b>	FPS	0.54	0.46
	Situation	Pos, below average	Pos
<b>sim card</b>	FPS	0.60	0.40
	Situation	Pos, below average	Pos, below average
<b>high speed</b>	FPS	0.29	0.71
	Situation	Pos, below average	Pos, below average
<b>food court</b>	FPS	0.27	0.73
	Situation	Pos, below average	Pos, below average
<b>souvenir shop</b>	FPS	0.62	0.38
	Situation	Pos, below average	Pos, below average
<b>rail bike</b>	FPS	1.00	-
	Situation	Pos, unique	-
<b>cable car</b>	FPS	1.00	-
	Situation	Pos, unique	-
<b>front desk</b>	FPS	0.29	0.71
	Situation	Pos, below average	Pos, below average
<b>hot spring</b>	FPS	-	1.00
	Situation	-	Pos, unique
<b>sky lantern</b>	FPS	0.43	0.57

	Situation	Pos, below average	Pos
<b>mrt station</b>	FPS	0.08	0.92
	Situation	Neu, below average	Pos, below average
<b>rock formation</b>	FPS	0.17	0.83
	Situation	Pos, below average	Pos, below average
<b>FPS AVG.</b>		0.51	0.52

可以從表 4-3、表 4-4、表 4-5、表 4-6 發現，當 $\sigma$ 從 0.1 增加至 0.4，會使高於市場平均的特徵減少，最後只剩下獨特的特徵，即市場上沒有其他應用有此特徵。

## 第四節 RE-SWOT 分析

根據 FPS 結果表，若 KKday 特徵為正向且高於市場平均水平，則定為 KKday 優勢；若 KKday 特徵為負向且低於市場平均水平，則定為 KKday 劣勢；若 Klook 特徵為正向且高於市場平均水平，則定為 KKday 威脅；若 Klook 特徵為負向且低於市場平均水平，則定為 KKday 機會。因本研究實驗了四種 $\sigma$ 值的結果，因此根據表 4-3 至表 4-6，分別產生表 4-7 至表 4-10 的 RE-SWOT 矩陣。

表 4-7 KKday 之 RE-SWOT 分析結果( $\sigma = 0.1$ )

$\sigma = 0.1$	KKday 特徵	Klook 特徵
正向(Pos)且 高於市場平均水平	S 優勢: tour guide observation deck pocket wifi credit card	T 威脅: high speed food court front desk hot spring

	qr code day tour souvenir shop rail bike cable car	mrt station rock formation
負向(Neg)且 低於市場平均水平	W 劣勢: sun moon lake	O 機會:

表 4-8 KKday 之 RE-SWOT 分析結果( $\sigma = 0.2$ )

$\sigma = 0.2$	KKday 特徵	Klook 特徵
正向(Pos)且 高於市場平均水平	S 優勢: tour guide pocket wifi credit card qr code day tour rail bike cable car	T 威脅: food court hot spring mrt station rock formation
負向(Neg)且 低於市場平均水平	W 劣勢: sun moon lake	O 機會:

表 4-9 KKday 之 RE-SWOT 分析結果( $\sigma = 0.3$ )

$\sigma = 0.3$	KKday 特徵	Klook 特徵
正向(Pos)且 高於市場平均水平	S 優勢: tour guide pocket wifi qr code	T 威脅: hot spring mrt station rock formation

	rail bike cable car	
負向(Neg)且 低於市場平均水平	W 劣勢: sun moon lake	O 機會:

表 4-10 KKday 之 RE-SWOT 分析結果( $\sigma = 0.4$ )

$\sigma = 0.4$	KKday 特徵	Klook 特徵
正向(Pos)且 高於市場平均水平	S 優勢: qr code rail bike cable car	T 威脅: hot spring
負向(Neg)且 低於市場平均水平	W 劣勢: sun moon lake	O 機會:

根據 RE-SWOT ( $\sigma = 0.1$ )的結果，我們可以看出 KKday 的優勢特徵有 tour guide(導遊), observation deck(觀景台), credit card(信用卡), qr code(二維條碼), day tour(日遊), sim card(電話卡), souvenir shop(禮物店), rail bike(軌道自行車), cable car(纜車)，共九項，應該保留或擴展其相關特徵。原始評論中與優勢特徵有關的敘述如下：與「導遊」有關的評論中，提及 KKday 活動導遊服務友善、行程規劃的很好、以及會詳細介紹當地景點故事背景等。與「觀景台」有關的評論提到 KKday 有方便且便宜的觀景台票券，並且遊客享受其中的景色，例如台北 101、上海中心大廈。提及信用卡的評論中表示平台使用線上信用卡支付，遊玩時不用在另外付費，使用平台配合的信用卡也有相關的優惠。與「二維條碼」有關的評論提及購買完成票券後，平台會給予二維條碼，不需再另外領票便於使用。與「日遊」相關評論提及平台提供多種地方、多時間段、包交通景點或是住宿等單日或多日活動。與「電話卡」

有關的評論則是提及可以在平台線上購買，外國旅客可以到旅遊當地機場領取，當下開通使用電話等服務。禮物店評論提及有些景點有禮物店可以購買禮物，也有可以在機場禮品店提取線上購入的禮品服務。與「軌道自行車」有關的評論多次提及南韓旅遊活動中的軌道自行車相當有趣。與「纜車」有關的評論提及纜車票券的購入方便，且價格優惠，以及景觀相當好，例如有臺灣貓纜、香港大嶼山等，這些特徵都是可以予以保留或是多加推出相關類型的服務。

KKday 的威脅有 high speed(高鐵), food court(美食), front desk(前台), hot spring(溫泉), mrt station(捷運), and rock formation(岩石)，共 6 項，KKday 可以針對這幾個特徵去效仿 Klook 相關特徵。依據原始評論中提及威脅特徵相關服務有，Klook 提供方便的線上高鐵票訂購服務，並且有折扣。與「美食」有關的評論提及景點附近有美食廣場的美食容易取得，例如台北 101、捷運美食街。與「前台」有關評論的指在 Klook 可以獲得比前台更好的服務或是票券折扣。與「溫泉」有關評論指其活動費用在 Klook 購買可以得到便宜的折扣。與「捷運站」有關提及 Klook 的旅遊團接駁地點鄰近捷運站。其中與「岩石」有關提及的是野柳的旅遊服務，包含接駁車票券等受到遊客的喜愛。這幾點 Klook 的服務將會是 KKday 的威脅，可以去效仿的部份。

KKday 劣勢中有 sun moon lake(日月潭)一項，應改善其相關產品或服務，依據評論中提劣勢特徵服務，多個評論提及 KKday 的日月潭遊的服務糟糕，集合地點的標示不清，出現導遊遲到或是出現問題但不退款等缺失應改善。

## 第五節 生成需求

基於 RE-SWOT 的結果，生成以下四種需求策略。

一、WT 策略：指在最小化弱點以使其不易受到威脅影響



的要求。

根據 KKday 的威脅特徵，KKday 可以降低或改善以下幾個類型的產品服務對遊客的影響，增加高鐵票券的線上販售及相關折扣、提供景點附近的美食廣場、改善購買線上票券的方便程度要比在前台便宜且方便、增加溫泉旅店的產品折扣及服務、效仿設置鄰近捷運站的接駁的點和提供野柳旅遊團的接駁及門票，以防止 KKday 用戶轉向 Klook 去尋求這些類型產品。

**二、ST 策略：指在利用優勢降低威脅脆弱性的要求。**

KKday 可以利用以下幾個類型的產品，利用推出更多優質的導遊、重視有觀景台的景點活動、利用信用卡折扣優惠、多加使用二維條碼提供票券、推出更多種類單多日的遊覽、電話卡的販售提取可以方便且多元化、禮物店可以增加更多不同的商品或是折扣服務、軌道自行車以及纜車活動帶動更多其他旅遊活動等，減低日月潭旅遊活動的不優對 KKday 帶來的威脅。

**三、WO 策略：指在克服劣勢以尋求機會的策略要求。**

KKday 可以改善自身日月潭旅遊活動的服務，可以更換配合的在地服務業者尋求其他服務機會。

**四、SO 策略：指在尋求契合優勢的機會的策略要求。**

KKday 可以加強提升其優勢並尋求其他機會。

其中因為 RE-SWOT 的結果缺少了機會的特徵，因此無法完全生成 WO 策略及 SO 策略。嘗試改變 $\sigma$ 值希望有機會的產生，但當 $\sigma$ 從 0.1 增加至 0.4，並沒有產生機會的特徵，反而減少了優勢及威脅高於市場平均的特徵，最後剩下的都是平台獨特的特徵。

## 第五章 結論

### 第一節 研究意涵

本研究的主要目的在於了解目的地旅遊體驗平台 KKday 和其競爭者 Klook 之間的競爭關係，並給予 KKday 建議改善其產品及服務。為此，我們利用自然語言處理結合情緒分析方式產生 SWOT 分析來了解 KKday 之競爭力。

因此，本研究主要貢獻是，利用爬取 TripAdvisor 中兩目的地旅遊體驗平台相關產品及服務的評論進行資料分析。將評論利用自然語言處理方式產生特徵詞組，結合 VADER 情緒分析工具產生評論的等級，來生成 RE-SWOT 矩陣，過去並未有過類似的作法的研究。

其次，本研究第二重要貢獻是，通過 RE-SWOT 生成需求。通過需求策略，KKday 可以找出改善方向以降低其外部威脅，增強其優勢以防劣勢對 KKday 的傷害。依據本研究的結果，建議 KKday 可以提供更多優質的導覽，推出更多種類一日或多日的體驗活動，包含有觀景台的景點、軌道自行車以及纜車等活動，票券販售多加使用二維條碼，電話卡的販售提取可以方便且多元化，搭配更好的信用卡折扣優惠，增加更多不同的商品或是折扣的禮物店服務，而日月潭旅遊活動是使用者較不滿意之處，建議 KKday 可再深入瞭解原因，以消除此項弱勢帶來的威脅。KKday 可以多加效仿 Klook 受好評的面向，例如：增加高鐵票券的線上販售及相關折扣、提供景點附近的美食廣場資訊、改善購買線上票券的方便程度、增加溫泉類型旅店的產品折扣及服務、設置鄰近捷運站的接駁的點，以防止 KKday 用戶轉向 Klook 去尋求這些類型產品。

第三，本研究以 KKday 為例，不過可以被使用於所有目的地旅遊體驗平台或是其他產業的公司，提供他們一個評估公司競爭力的方式。公司可以針對線上的評論了解到顧客的主要需求以及

自身公司和競爭對手公司的優缺點，不用再另外耗費時間和成本收集問卷，亦可以作為市場調查的一部分。本研究競爭對手只用 Klook，但是 RE-SWOT 分析不限制於只針對其中一個競爭者。如若企業公司想要使用 RE-SWOT 分析來了解競爭力，建議收集的評論來自於評論數較多、公開且公平的第三方平台。分析完成後企業可以針對優勢和劣勢，以及對手給予的機會和威脅，做出產品策略或是發展策略上調整。

第四在 RE-SWOT 分析過程中有幾個需要注意及調整的地方。首先，本研究使用的情緒分析方法需要經過轉換才能使用於分析中，在結果中我們發現經過轉換會缺少極正與極負的等級，所以如何決定轉換極性強度分類規則時的門檻值，是需要再研究改善的問題。雖然 RE-SWOT 分析過程中需要用使用情緒分析方式來代替缺少的等級評分，但是因轉換過程中門檻值的問題，所以也可以考慮使用其他情緒分析方式來產生等級評分。再則，競爭對手的多少會影響 FPS 的值，所以公式中  $\sigma$  的設定值，會需要針對需要做一些調整。最後，經過詞組搭配產生的詞組特徵，可以用來計算評論中所要表達的特徵，但是生成需求時無法直接做解讀，需要以原評論中的提及的內容來解釋其需求。

最後，研究發現旅客對於目的地旅遊體驗平台的需求主要有三點，購買的票卷或旅遊產品便於使用、體驗活動的服務和類型以及價格的折扣。對於平台要推出新產品時因該先考慮以下三個問題：第一，線上購買的票券或旅遊產品是否可以比當場購買方便且便於使用？第二，目的地體驗活動的服務是不是符合遊客喜愛景點及有良好服務的期待？第三，相同類型的產品在其他目的地體驗平台中使否用有更好的優惠折扣？

## 第二節 研究限制

### 一、資料來源

由於本研究只取用 TripAdvisor 中的 KKday 及 Klook 英文評論。TripAdvisor 為旅遊網站平台內容廣泛，並不是專門針對目的地旅遊體驗平台的評論，所以爬取的評論在經過處理之後，評論文本數並不是很多。

### 二、使用者評論的語系

雖然 TripAdvisor 有不同語系的評論，但是本研究以英語系的評論為主。

因本研究的評論文本為英文，但撰寫的方式為中文，需要經過翻譯的過程，因此部分需求和特徵的英文意涵，無法精準的以中文表達。

### 三、 $\sigma$ 設定

本研究分別以  $\sigma = 0.1, 0.2, 0.3, 0.4$  作為公式的設定值，來計算 FPS。從本研究的實驗結果可以看出，不同的  $\sigma$  值會得到不同的 RE-SWOT 矩陣，本研究在此未探討如何決定合適的  $\sigma$  值。

## 第三節 未來展望

本研究在發展和概念驗證方面都為未來的研究提供了許多機會。實際上，有必要進行更多研究以完善和進一步完善我們的發現。

首先，未來研究可以在取得更多不同旅遊平台中 KKday 的相關評論，增加其文本數量，可以更精確的了解其他的產品及服務評價。也可以增加其他國家的目的地旅遊體驗平台作為競爭者進

行比較，了解全球目的地旅遊體驗平台的發展，以及外國旅客的需求等。

其次，未來研究可以進一步改善 VADER 的情緒分數在轉換極性強度分類規則時的門檻值，或是使用其他不同情緒分析方法去產生等級分數。

最後，未來 RE-SWOT 分析研究，只要有顧客的評論皆可以被使用於飯店、集團、航空業或是其他不同的產業的業者，進行競爭力分析方法。



## 參考文獻

### 一、中文部分

- Anny. (2018), KKday 獲千萬美元投資！日本最大國際旅遊集團 H.I.S.攜手優化旅客體驗[線上資料]，來源：  
<https://www.inside.com.tw/article/12052-kkday-b-round-from-japan-travel-company-his> [2019, December 20]
- Frankie. (2016), 七成國人旅遊愛民宿「亞洲萬里通」攜手 Tripitta 國內旅遊哩程加倍送[線上資料]，來源：  
<https://www.gq.com.tw/life/content-29191> [2019, December 20]
- Huang C.H. (2016), Data Cleaning 介紹[線上資料]，來源：  
<https://medium.com/@chihsuan/data-cleaning-介紹-387d24666b1f> [2020, March 18]
- Huang, Y. (2018), NLTK 初學指南(二)：由外而內，從語料庫到字詞拆解 — 上手篇[線上資料]，來源：  
<https://medium.com/pyladies-taiwan/nltk-初學指南-二-由外而內-從語料庫到字詞拆解-上手篇-e9c632d2b16a> [2020, March 18]
- INSIDE 硬塞的網路趨勢觀察(2016), 台灣旅遊電商 KKday 2.2 億台幣 A+輪投資，將於亞洲 10 國設中心[線上資料]，來源：  
<https://www.inside.com.tw/article/7832-kkday-220-m> [2019, December 20]

INSIDE 硬塞的網路趨勢觀察(2018)，KKday 宣佈獲得 LINE Ventures 的 B+輪投資[線上資料]，來源：  
<https://www.inside.com.tw/article/14650-kkday-series-b-line-ventures> [2019, December 20]

KATI (2019)，JOT DOWN VOYAGE OTA 時代來臨：客路 KLOOK vs KKday 是誰？哪個評價比較好？[線上資料]，來源：  
<https://jotdownvoyage.com/klook-vs-kkday/#OTAOTA> [2019, December 20]

KKday (2019)，[線上資料]，來源：  
<https://event.kkday.com/zh-tw/about/about> [2019, December 20]

Klook (2019)，KLOOK 獲日本軟銀願景基金 2.25 億美元 D+輪融資，D 輪融資累計達 4.25 億美元[線上資料]，來源：  
<https://www.klook.com/zh-TW/newsroom/content/6405?n=0> [2019, December 20]

Lien, L. (2020)，SWOT 分析這樣做，寫出成功的經營策略[線上資料]，來源：  
<https://www.hububble.co/blog/swot/> [2020, May 5]

TechNews (2018)，KKday 宣布獲得 LINE Ventures 投資，策略合作加速旅遊數位化[線上資料]，來源：  
<https://technews.tw/2018/11/08/kkday-line-ventures/> [2019, December 20]

TripAdvisor (2019), 關於 Tripadvisor [線上資料], 來源:  
<https://tripadvisor.mediaroom.com/tw-about-us> [2019,  
December 20]

呂尉慈, 杜淑芬(2020), 應用文字探勘技術於臺灣體驗行程評價  
之情緒分析, 發表於智慧觀光旅遊新創研討會, 臺北市:  
中國文化大學。

何佩珊(2018), 旅遊體驗平台雙雄之爭: KLOOK 能用雄厚資金  
擠下對手 KKday 嗎? [線上資料], 來源:  
<https://www.bnext.com.tw/article/50160/travel-experience-competition-of-klook-and-kkday> [2019, December 25]

何佩珊(2018), LINE 入股 KKday, 還找來雄獅、Agoda 合攻旅  
遊市場 [線上資料], 來源:  
<https://www.bnext.com.tw/article/51215/line-travel-kkday-liontravel-agoda-tripresso> [2019, December 20]

吳元熙(2017), 獲紅衫資本投資 3,000 萬美元, 行程預訂平台  
KLOOK 正式進軍台灣 [線上資料], 來源:  
<https://www.bnext.com.tw/article/43390/klook-completes-us-dollars-30millions-series-b-funding> [2019, December 20]

吳元熙(2017), 旅遊電商 KLOOK 獲 18 億 C 輪融資, 將砸 5 億  
投入台灣自由行市場 [線上資料], 來源:  
<https://www.bnext.com.tw/article/46697/klook-raised60-million-dollars-series-c-funding> [2019, December 20]



吳元熙(2018)，旅遊平台 KKday 獲 3 億銀彈，聯手日本旅遊集團 H.I.S.打國際戰[線上資料]，來源：

<https://www.bnext.com.tw/article/48257/taiwanese-tourism-startup-kkday-closed-series-b-ten-million-dollars-funding>  
[2019, December 20]

呂維苓，黃仕斌(2016)，體驗經濟下的產業轉型-以旅遊電商 KKday 為例，國立交通大學企業管理碩士學程論文，新竹市：國立交通大學。

李佳孟(2019)，應用深度學習於社群媒體情緒分析-以軍事類網路留言版為例，國防大學管理學院資訊管理學系碩士論文，臺北市：國防大學管理學院。

明報財經(2017)，港產 Startup 獲基金青睞 local tour 訂購平台 Klook 融資 2.3 億[線上資料]，來源：

<https://www.mpfinance.com/fin/instantf2.php?node=1488437371765&issue=20170302> [2019, December 20]

亮(2016) 旅遊產業價值鏈[線上資料]，來源：

<https://www.stockfeel.com.tw/從旅行社看產業價值鏈/>  
[2019, December 10]

侯姿伶(2018)，創新服務對消費者價值及服務體驗的影響-以旅遊電商 KKday 為例，臺中科技大學企業管理系碩士班學位論文，臺中市：臺中科技大學。

姚舜(2015)，KKday.com 旅遊網開站 帶客赴義體驗神鬼戰士[線上資料]，來源：

<https://www.chinatimes.com/realtimenews/20150113002027-260405?chdtv> [2019, December 20]

姚舜(2015)，KKday 募試玩員 日、韓爽玩 30 天[線上資料]，來源：

<https://www.chinatimes.com/realtimenews/20150331004839-260405?chdtv> [2019, December 20]

徐雅琪(2016)，目的地旅遊平台之創新商業模式—以 KKday.com 為例，國立中山大學資訊管理學系研究所學位論文，高雄市：國立中山大學。

張文騰(2017)，旅遊體驗平台之創新商業模式-以 Klook 為例，國立臺灣師範大學全球經營與策略研究所學位論文，臺北市：國立臺灣師範大學。

張旭宏(2015)，Klook 客路 APP 獲 A 輪融資 1.5 億用於深耕台灣市場[線上資料]，來源：

<https://news.cnyes.com/news/id/315277> [2019, December 20]

張傳珩(2017)，文本探勘與情緒分析於產品推薦之應用-以 PTT 電影版為例，東吳大學商學院資訊管理學系碩士論文，臺北市：東吳大學。

莊士賢(2018)，KKday、卡通 Poli 聯名巴士小朋友就愛它！[線

上資料]，來源：

<https://ent.ltn.com.tw/news/breakingnews/2465474> [2019, December 20]

郭炳良(2017)，應用文字探勘技術分析電影評論與台灣票房之關係，中華大學生物資訊學系碩士論文，新竹市：中華大學。

陳君毅(2018)旅遊預定平台 KLOOK 獲 60 億 D 輪融資，將加速拓展歐美市場[線上資料]，來源：

<https://www.bnext.com.tw/article/50153/klook-d-round>  
[2019, December 20]

陳君毅(2019)，KKday 宣佈獲 Cool Japan Fund 3 億台幣投資，日本政府也是大股東！[線上資料]，來源：

<https://www.bnext.com.tw/article/54572/kkday-funding-2019-august> [2019, December 20]

游綉雯(2015)，使用情緒分析於社群論壇消費者評論滿意度評估之研究-以 TripAdvisor 旅遊網站為例，國立中興大學行銷學系碩士論文，臺中市：國立中興大學。

楊孟芯(2019)，出國行程排好了嗎？你用 KKDAY 還是 KLOOK？線上旅遊大餅，從孫正義到 GOOGLE 都在搶[線上資料]，來源：<https://www.storm.mg/article/1550934>  
[2019, December 20]

楊倩蓉(2016)，Klook 客路旅遊體驗，用科技搞定玩樂大小事[線上資料]，來源：

<https://web.archive.org/web/20160821221108/http://www.cna.com.tw/magazine/20/201608090002-1.aspx> [2019, December 20]

極客公園(2015)，客路旅行 Klook 獲騰訊高管領投 150 萬美元種子融資[線上資料]，來源：

<https://kknews.cc/travel/vpa8lll.html> [2019, December 20]

葉怡君(2016, December 5)，KKday A+ 輪融資獲 2.2 億資金[線上資料]，來源：

<http://investtw.net/index.php?c=news&uid=146828> [2019, December 20]

韓勇(2006)，旅行社經營管理，大陸：北京大學出版社。

## 二、英文部分

Ahani, A., Nilashi, M., Ibrahim, O., Sanzogni, L., & Weaven, S.

(2019). Market segmentation and travel choice prediction in Spa hotels through TripAdvisor's online reviews.

*International Journal of Hospitality Management*, 80, 52-77.

Aksu, A., & BAYAR, K. (2019). Development of Health Tourism in Turkey: SWOT Analysis of Antalya Province. *Journal of Tourism Management Research*, 6(2), 134-154.

Alaei, A. R., Becken, S., & Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. *Journal of Travel Research*, 58(2), 175-191.

Chowdhury, G. (2003) Natural language processing. *Annual Review of Information Science and Technology*, 37, 51-89.

Dalpiaz, F., & Parente, M. (2019). RE-SWOT: from user feedback to requirements via competitor analysis. In *International Working Conference on Requirements Engineering: Foundation for Software Quality* (pp. 55-70). Springer, Cham.

Dunning, T. E. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational linguistics*, 19(1), 61-74.

Elbagir, S., & Yang, J. (2019). Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment. In *Proceedings of the International MultiConference of Engineers and Computer Scientists* (pp. 122-16).

Fu, Y., Hao, J. X., Li, X., & Hsu, C. H. (2019). Predictive accuracy of sentiment analytics for tourism: a metalearning perspective on Chinese travel news. *Journal of Travel Research*, 58(4), 666-679.

Hou, Z., Cui, F., Meng, Y., Lian, T., & Yu, C. (2019). Opinion mining from online travel reviews: A comparative analysis of

Chinese major OTAs using semantic association analysis. *Tourism Management*, 74, 276-289.

Hutto, C. J., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

Ivanov, S. (2019). Modeling Company Sales Based on the Use of SWOT Analysis and Ishikawa Charts. In *CEUR Workshop Proceedings* (Vol. 2422, pp. 385-394).

Kamran, M., Fazal, M. R., & Mudassar, M. (2020). Towards empowerment of the renewable energy sector in Pakistan for sustainable energy evolution: SWOT analysis. *Renewable Energy*, 146, 543-558.

Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X. (2018). Automated sentiment analysis in tourism: Comparison of approaches. *Journal of Travel Research*, 57(8), 1012-1025.

Kuhamanee, T., Talmongkol, N., Chaisuriyakul, K., San-Um, W., Pongpisuttinun, N., & Pongyupinpanich, S. (2017). Sentiment analysis of foreign tourists to Bangkok using data mining through online social network. In *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)* (pp. 1068-1073). IEEE.

MarketingTeacher. (2017). *SWOT Examples* [Online]. Available:  
<http://www.marketingteacher.com/swot-analysis-examples/>  
[2020, June 25]

Masrury, R. A., & Alamsyah, A. (2019). Analyzing Tourism Mobile Applications Perceived Quality using Sentiment Analysis and Topic Modeling. In *2019 7th International Conference on Information and Communication Technology (ICoICT)* (pp. 1-6). IEEE.

Maulani A.M.A. (2016). *KKday secures US\$4.5M Series A led by AppWorks Ventures* [Online]. Available: <https://e27.co/fun-begins-kkday-secures-us4-5m-series-led-appworks-ventures-20160316/> [2020, June 25]

Newman, H., & Joyner, D. (2018). Sentiment analysis of student evaluations of teaching. In *International conference on artificial intelligence in education* (pp. 246-250). Springer, Cham.

Niranjanamurthy, M., Nithya, B. N., & Jagannatha, S. (2019). Analysis of Blockchain technology: pros, cons and SWOT. *Cluster Computing*, 22(6), 14743-14757.

Owler (2019). *Kkday's Competitors, Revenue, Number of Employees, Funding and Acquisitions* [Online]. Available:  
<https://www.owler.com/company/kkday> [2019, December 25]

Pandey, P. (2018). *Simplifying Sentiment Analysis using VADER in Python (on Social Media Text)* [Online]. Available: <https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f> [2019, December 20]

Parente, M. G. (2018). Using NLP and Information Visualization to analyze app reviews (Master thesis) [Online]. Available: <https://dspace.library.uu.nl/handle/1874/368082> [2020, May 30]

*Part of Speech Tagging with NLTK*. (n. d. ). [Online]. Available: <https://pythonprogramming.net/part-of-speech-tagging-nltk-tutorial/> [2020, July 25]

Prameswari, P., Surjandari, I., & Laoh, E. (2017). Mining online reviews in Indonesia's priority tourist destinations using sentiment analysis and text summarization approach. In *2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)* (pp. 121-126). IEEE.

Qi, J., Zhang, Z., Jeon, S., & Zhou, Y. (2016). Mining customer requirements from online reviews: A product improvement perspective. *Information & Management*, 53(8), 951-963.

Ramanathan, V., & Meyyappan, T. (2019). Twitter text mining for sentiment analysis on people's feedback about oman tourism.



In *2019 4th MEC International Conference on Big Data and Smart City (ICBDSC)* (pp. 1-5). IEEE.

*Sentiment Analysis*. (n. d. ). [Online]. Available:

<https://web.stanford.edu/class/cs124/lec/sentiment.pdf> [2020, July 25]

Swarnkar, N. (2020). VADER Sentiment Analysis in Algorithmic Trading[Online]. Available: <https://blog.quantinsti.com/vader-sentiment/> [2020, June 18]

Thanopoulos, A., Fakotakis, N., & Kokkinakis, G. (2002). Comparative Evaluation of Collocation Extraction Metrics. In *LREC* (Vol. 2, pp. 620-625).

Tsuruoka, Y., Tateishi, Y., Kim, J. D., Ohta, T., McNaught, J., Ananiadou, S., & Tsujii, J. I. (2005). Developing a robust part-of-speech tagger for biomedical text. In *Panhellenic Conference on Informatics* (pp. 382-392). Springer, Berlin, Heidelberg.

Vijayarani, S., & Janani, R. (2016). Text mining: open source tokenization tools-an analysis. *Advanced Computational Intelligence: An International Journal (ACIJ)*, 3(1), 37-47.

Wang, J., & Wang, Z. (2020). Strengths, weaknesses, opportunities and threats (Swot) analysis of china's prevention and control

strategy for the covid-19 epidemic. *International Journal of Environmental Research and Public Health*, 17(7), 2235.

Webster, J. J., & Kit, C. (1992). Tokenization as the initial phase in NLP. In *COLING 1992 Volume 4: The 15th International Conference on Computational Linguistics*.

Weihrich, H. (1982). The TOWS matrix—A tool for situational analysis. *Long range planning*, 15(2), 54-66.

Zhao, Y., Xu, X., & Wang, M. (2019). Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International Journal of Hospitality Management*, 76, 111-121.

