

# 巴黎奧運對當地居民情緒與經濟影響之分析

## Abstract

本研究探討 2024 年巴黎奧運對當地居民在經濟與情緒層面所造成的影響，聚焦於奧運前後房價變化與社群媒體情緒波動之間的潛在關聯。我們蒐集了巴黎 Airbnb 房價數據，以及社群媒體 Reddit 上的相關留言，利用自然語言處理技術進行情緒分數分析，並透過多種監督式回歸模型，檢驗價格特徵對社群情緒的預測能力。

研究結果發現，無論是在特徵工程前或是將留言分類後的模型訓練中，各模型的  $R^2$  分數皆極低，顯示留言情緒與價格特徵間並無關聯。

雖然模型預測效果有限，但研究的價值在於：嘗試把「情緒」和「經濟」兩種資料結合，從跨領域的視角，觀察大型活動對當地居民情緒和生活的影響。此外，也提供了一些資料處理或建模方面的改善方向。

## 1. Introduction 簡介

### 1.1 研究背景

奧運作為全球矚目的體育賽事，對主辦城市而言，不僅涉及經濟發展，更牽動居民的生活成本與情緒反應。在觀光旅遊與基礎建設成長等經濟效益背後，可能伴隨著短期租金上漲、生活成本增加等負面影響，進而引發居民情緒波動之問題，特別是在短期住宿市場的租金變化劇烈，對民眾生活壓力與社會觀感可能產生深遠影響。

### 1.2 研究動機

過往研究多著重於奧運對經濟層面的效益評估，如觀光旅遊、基礎建設或就業等，但較少有文獻深入探討「經濟變化是否會反映在居民的情緒上」。我們認為，情緒與經濟之間存在潛在的互動關係，尤其是在生活成本的壓力增加下，居民可能產生不滿或是負面情緒。

因此，我們希望探索巴黎奧運前後，當地居民的情緒與經濟變動之間是否存在關聯，並進一步評估是否能以經濟變數來預測社群情緒，彌補現有研究的不足，使情緒數據成為補充經濟分析的其中一個關鍵來源。

### 1.3 問題陳述

透過釐清以下問題，理解大型活動對居民生活的影響：

- 短期住房價格變化與社群情緒之間是否有顯著關聯
- 是否能透過機器學習模型，建立從經濟指標去預測情緒的方法

希望基於研究結果，提出改善城市居民經濟福祉的建議，並針對情緒負面影響提出緩解策略，為未來奧運主辦城市提供經濟與情緒管理的最佳實踐。

## 2. Related Work 相關文獻探討

關於奧運影響評估與社群媒體情緒分析的文獻橫跨多個學科領域，從經濟學、都市規劃到計算語言學與社會心理學。本文獻回顧綜合三個主要研究領域的相關研究：(1)大型活動對主辦城市的經濟影響、(2)社群媒體資料的情緒分析方法論、(3)經濟因素與公眾情緒的交互關係。

### 2.1 奧運對主辦城市的經濟影響

關於主辦奧運經濟效應的研究在過去數十年間有了顯著發展，尤其著重於其承諾效益與實證分析。Baade 和 Matheson (2016) 提供奧運成本效益的全面性分析，其結論為在大多數情況下，奧運對主辦城市而言是虧損的投資；只有在非常特定的情況下才會產生正面的淨效益。此發現挑戰奧運組織者所宣傳的傳統敘述，並與對大型活動影響的廣泛質疑一致。

外交關係委員會 (Council on Foreign Relations, 2024) 報告指出，主辦奧運的成本已急劇上升，而經濟效益卻遠不明確，並提及北京 2008 年夏季奧運產生 36 億美元的收入，相較於超過 400 億美元之成本，凸顯主辦此類活動所面臨的財務風險。

#### 2.1.1 住房市場效應

房地產市場代表奧運影響的關鍵面向，其直接影響當地居民的經濟福祉。Kontokosta (2012) 進行奧運對住宅房地產市場影響最嚴謹的分析之一，利用 1984 年至 2000 年間六個主辦城市的資料，並採用調整後的中斷時間序列方法，研究發現奧運並非一體適用的經濟發展策略，潛在結果取決於多項因素，包括規劃協調、奧運相關發展以及總體奧運投資的相對規模。

關於 2012 年倫敦奧運的研究發現，主辦行政區的房地產售價提高了 2.1% 至 3.3%，尤其取決於影響區域的定義，距離主要奧運場館三英里內的房地產售價提高了 5%，對主辦行政區房價的整體影響達 14 億英鎊。2024 年巴黎奧運提供住房市場動態的案例研究，對度假租賃市場的分析發現奧運期間住宿的平均每晚價格比 2023 年同期高出五倍，從 141 歐元上升至 706 歐元。

#### 2.1.2 就業影響效應

奧運的就業影響往往未達預期，以鹽湖城主辦 2002 年奧運為例，該市僅增加 7,000 個工作崗位，約為官方預測數量的 10%，且大多數工作都給原本就擁有工作的人員，意味著失業統計數據並未顯著改善。尤其使用異質性穩健事件研究分析的最新研究 (Billings & Holladay, 2012; Maennig & Richter, 2012; Langer et al., 2018) 顯示，奧運影響的結果好壞參半，尤其早期研究中發現的正面效應往往在為奧運主辦國建構適當對照組後消失。

### 2.2 社群媒體情緒分析研究

情緒分析領域已從基本的極性檢測，發展為結合時間動態、網路關係和情緒強度檢測的複雜多模態分析。情緒分析應用可見於國際事件、醫療保健、政治和商業等 (Talha et al., 2019)，大多數文章採用意見詞典法分析社群媒體中的文本情緒，且主要從微網

誌網站 **Twitter** 提取資料。近年情緒分析研究，強調自然語言處理中評估和識別文本資料中傳達的情緒語調或情感之方法，其重要性在於能從大量文本資料中獲得有價值的洞察，掌握使用者情感與改善決策 (Zhang et al., 2024)。此外，關於情緒分析的當前挑戰包含有效提取使用者從大量評論中的態度，特別是在基於網路應用程式的擴展導致評論量爆炸性增長的情況下 (Xu et al., 2022)。

## 2.3 經濟因素與公眾情緒的交集

雖然奧運經濟影響和社群媒體情緒分析分別都有大量研究，然而，目前很少有研究試圖在大型活動期間建立經濟變數與公眾情緒之間的直接聯繫。

### 2.3.1 理論框架

情緒分析與金融應用之跨領域研究發現，自彭博社開創性工作提出新聞情緒對市場波動性的影響，情緒分析已廣為運用 (Puschmann & Powell, 2018)。此外，關於情緒分析作為理解使用者之研究強調，情緒分數能作為評估社群媒體使用者情感的客觀手段 (Puschmann & Powell, 2018)。

然而，經濟與情緒資料的整合面臨方法論挑戰，經濟領域之系統性回顧發現社群媒體情緒分析中，與資料集、文本語言、分析方法和評估指標相關的跨領域研究，在試圖與經濟變數建立關聯時，將變得更加複雜 (Xu et al., 2022)，雖然情緒分析已被證明是在不同學科中衡量公眾意見的有價值工具，且已成功應用於金融市場、政治、經濟和健康領域，但其應用仍在各領域中分佈不均且面臨不同困難 (Song et al., 2023)。

## 2.4 研究缺口與貢獻

儘管在奧運經濟學和情緒分析方面已有大量研究，但在理解兩者交集方面仍存在重大空白。大多數奧運影響研究關注總體經濟指標，而非檢視經濟變化如何轉化為居民的公眾情緒和情感反應。此外，雖然情緒分析已應用於各種危機和事件情境，但其在奧運等大型活動中的應用仍有限，尤其是檢視住房市場動態與居民情緒之關係。

本研究主要貢獻包含：

- (1) 在奧運影響評估中開創性整合經濟 (Airbnb 價格) 和情緒 (Reddit 評論) 資料
- (2) 應用機器學習技術從經濟指標預測情緒
- (3) 提供關於大型活動期間短期住房市場波動與公眾情緒之間關係的實證證據。

## 3. Data Description 資料說明

### 3.1 資料來源

- 情緒資料：蒐集自 Reddit (包含 france、paris 兩個 subreddit)，透過關鍵字如 “Airbnb”、“Olympics”、“生活品質”等篩選。
- 經濟資料：透過Kaggle平台，蒐集Airbnb之價格資料，涵蓋巴黎地區 2023-2025 年間的日租房價。

### 3.2 資料數量

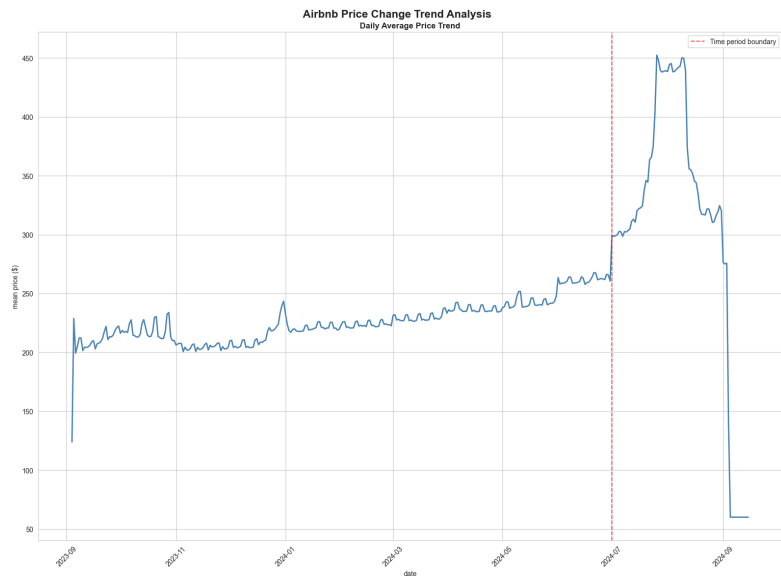
reddit評論	奧運前	奧運後	總和
airbnb相關	7835	3100	10935
olympics相關	21075	13425	34500

### 3.3 資料前處理過程

#### 3.3.1 Reddit

- 留言中涉及「奧運」、「生活成本」、「Airbnb」等主題做分類與優先排序
- 使用 Google Translate API 進行非英文留言的翻譯(法文→英文)
- 使用 TextBlob 進行英文留言的情緒分數分析(Polarity -1 到 +1)

#### 3.3.2 Airbnb



圖一 2023年9月至2024年9月間巴黎地區 Airbnb 每日平均房價的變化

註:造成數值大幅落差的原因之一是部分房型在特定時間點缺乏資料(如2023-09-05或2024-09-06)

- 以 7/1 為分界(圖一中, 7/1為第一次驟升點)分為奧運前後
- 接著整理出以下, 作為全面特徵:

avg\_price-平均價格

avg\_price\_change-平均價格變化

median\_price-價格中位數

std\_price\_change-價格變化的標準差

price\_range-價格範圍(最大值-最小值)

log\_avg\_price-平均價格對數

relative\_price\_change-相對價格變化百分比

## 4. Methodology 研究方法

本研究旨在探討 Airbnb 價格在奧運前後是否與社群情緒變動存在可預測性關聯。目標變數為由 TextBlob 分析出的留言情緒分數。特徵則為從 Airbnb 價格資料中處理過的數據，如價格變化、分散程度與時間區分等，作為模型學習預測依據。

由於社群情緒受眾多潛在因素影響，資料極具雜訊，因此在實作層面，我們將資料依照留言主題 (Airbnb、Olympic) 分別建立模型，統一採用多種監督式回歸模型，並搭配特徵工程與模型評估流程。

### 4.1 模型選擇

採用多種監督式回歸模型，以驗證 Airbnb 房價特徵對情緒的預測效果

- 線性模型: Linear Regression、Ridge、Lasso、ElasticNet
- 樹模型: Decision Tree、Random Forest、Gradient Boosting、AdaBoost
- 其他模型: K-Nearest Neighbors (KNN)、Support Vector Machine (SVM)、XGBoost、LightGBM

### 4.2 實作流程與特徵工程

- 以airbnb價格特徵(共8個)作為自變數，情緒分數作為目標變數
- 模型評估指標為 $R^2$
- 執行特徵重要性分析，以解釋預測情緒變化的重要價格特徵

特徵工程前的訓練結果顯示，將未經處理的情緒資料進行訓練， $R^2$ 最高只有0.0007，模型僅能解釋0.1%的情緒變異，無法有效預測留言情緒。

因此，我們進一步實作特徵工程，將留言分類為 Airbnb相關或 Olympic相關，分別建立模型，預期可以提高相關性。

### 4.3 使用工具與技術架構

初期預期使用autosklearn，但在多種環境下安裝受限 (Windows/Colab)，且使用使用wsl安裝完成後，出現記憶體不足之問題，最後改為手動配置。

- 語言: Python
- 使用套件: 主要使用TextBlob (針對英文留言做情緒分數分析)、pandas、sklearn、xgboost等工具。

## 5. Results & Analysis 實驗結果與分析

### 5.1 模型整體表現分析

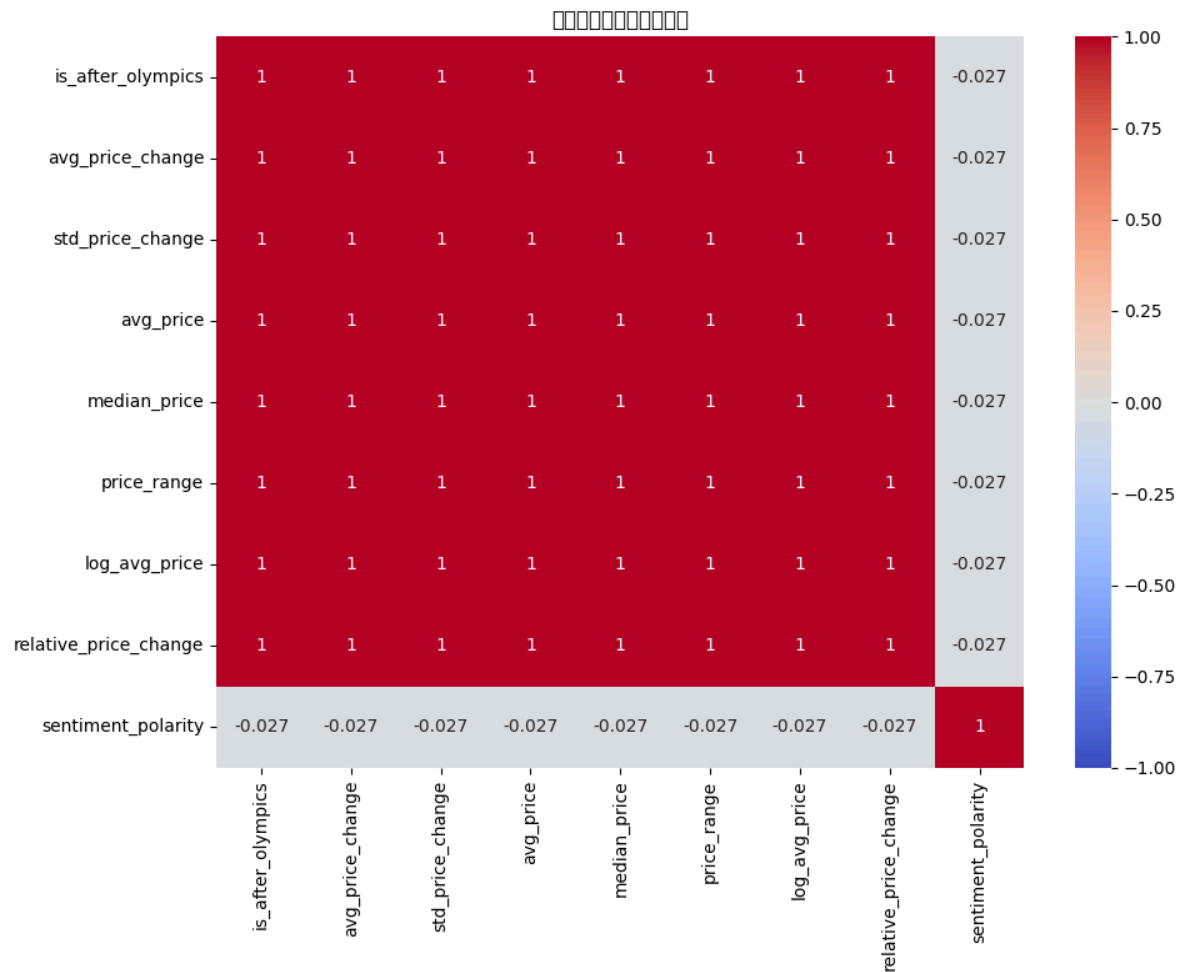
為驗證 Airbnb 價格是否能有效預測社群情緒，我們使用 TextBlob 對 Reddit 留言進行情緒分數分析，並分別以特徵工程前、Airbnb 相關留言與 Olympic 相關留言進行模型訓練與  $R^2$  評估。

Model	特徵工程前 $R^2$	Airbnb相關 $R^2$	Olympic相關 $R^2$
Linear Regression	0.0006	0.0021	-0.0003
Ridge Regression	0.0007	0.0018	-0.0002
Lasso Regression	-0.0002	-0.0000	-0.0002
ElasticNet	-0.0002	-0.0000	-0.0002
Decision Tree	0.0007	0.0018	-0.0002
Random Forest	0.0007	0.0018	-0.0003
Gradient Boosting	0.0007	0.0018	-0.0002
AdaBoost	-0.0018	0.0020	-0.0004
KNN	-0.0061	-0.7025	-0.0607
SVM	-0.0454	-0.0043	-0.0092
XGBoost	0.0007	0.0018	-0.0002
LightGBM	0.0007	0.0018	-0.0002

從模型效能結果可觀察以下現象：

- 所有模型的  $R^2$  皆極低，接近於0，幾乎無預測能力
- 即使經過特徵工程處理，Airbnb 相關留言中的最佳模型(Linear Regression)  $R^2$  也僅提升至 0.0021
- Olympic 相關留言的模型，沒有任何一個呈現正向  $R^2$ ，甚至出現負值，顯示模型表現極差。

5.2 視覺化分析



圖二: Airbnb 價格特徵與情緒分數之間的相關性熱圖

利用相關性熱圖視覺化特徵與情緒的關係，發現兩者並沒有關係

6. Discussion 討論

6.1 結果意義與反思

透過熱圖分析發現，社群的情緒與房價變化之間缺乏顯著的關聯，留言情緒與所有價格特徵的相關係數皆接近 0，導致迴歸模型預測力有限。這可能代表兩者之間本就沒有明確可預測的關係，或是變數間的關係並非線性。也可能反映出，社群上的情緒表現，不是單靠價格變化這類數據就能解釋，而是更受到留言當時的語氣、或是使用者本身的想法與價值觀所影響。



## 6.2 資料與模型的限制

- 資料來源與樣本限制:主要採用 **Reddit** 的貼文與留言作為情緒來源,以「奧運」、「巴黎」、「房價」、「**Airbnb**」等關鍵字篩選資料。忽略了使用者未必能代表巴黎當地居民的觀點。篩選條件同時限制了留言的數量與多樣性,使資料筆數有限。
- 情緒分析工具限制:使用 **TextBlob** 進行情緒分數的標記,在處理非正式語言(如 **Airbnb** 留言)時的精確度有限,且不支援多語言情境下的細膩語義理解,可能導致情緒分數無法有效反映使用者真實情緒。
- 特徵工程的不足:雖嘗試建立多種價格變化統計特徵(如平均變化、標準差、中位數、對數處理等),但這些數值變數未能充分捕捉與情緒波動相關的語境或事件特徵,導致預測力有限。也有可能是因為使用的變數偏向一般經濟面,缺乏與情緒直接相關的內容變數。
- 建模方向與挑戰:目前設計為以價格特徵預測情緒分數的回歸問題,但實際上情緒通常是屬於分類性質(如正/負/中),可能導致模型不易找出適當關係。

## 7. Conclusion & Future Work 結論與未來工作

### 7.1 研究總結

本研究的目的是探討大型國際活動對城市居民情緒的潛在影響,原先假設 **Airbnb** 價格變化與社群情緒之間可能存在可觀察的關聯,尤其在重大事件如奧運前後。

然而,實際建模結果卻顯示這樣的預測力極其有限。雖然我們嘗試透過特徵工程處理,以及各類回歸模型,但情緒的複雜性顯然超出了單一數據可解釋的範圍。

從資料蒐集、特徵設計,到模型實作與評估,每一步都揭示了科學在處理主觀情緒與心理現象時所面臨的困難與挑戰。但儘管如此,特徵工程的嘗試仍具啟發性,雖然模型效能提升有限,但未來建模時若加強特徵設計,可能將有助於模型捕捉到關係。

### 7.2 未來工作與改進方向

- 將情緒分數改為分類形式:將連續的情緒分數轉換成「三分類」或「二分類」,重新改用分類模型,可能比回歸模型更能學習出模式。
- 使用進階情緒分析模型:如 **VADER**、**BERT**、**RoBERTa** 等更適合處理非正式語言與多語背景的工具。
- 補強特徵設計:增加如留言主題分類、留言熱度、發文時間是否落在重大新聞事件當日、地理位置資訊等特徵。



## Acknowledgements 致謝

本專題能夠順利完成，首先要感謝何承遠老師在研究過程中提供的指導與建議，在遇到問題與困難時給予實質上的協助與方向上的啟發。

感謝 Reddit 與 Airbnb 平台所提供的開放性資料，讓我們得以取用留言資訊與經濟活動資料，作為本研究的分析基礎。

也特別感謝組內成員在不同階段所付出的努力與合作：張逸安同學負責資料爬蟲與前置處理、黃裕媃同學負責相關文獻的探討與情緒分數分析，周聖詠同學主導機器學習與模型建構、謝蕙宇同學負責整合內容與撰寫報告，大家分工明確，讓整體進度能穩定推進。

最後，也感謝課程提供這個實作機會，讓我們有機會將課堂所學，實際應用於社會議題的分析中，培養跨領域整合的能力。

## References 參考文獻

- Baade, R. A., & Matheson, V. A. (2016). Going for the gold: The economics of the Olympics. *Journal of Economic Perspectives*, 30(2), 201-218. <https://doi.org/10.1257/jep.30.2.201>
- Firgo, M. (2021). The causal economic effects of Olympic Games on host regions. *Regional Studies*, 55(5), 903-914. <https://doi.org/10.1080/00343404.2020.1847270>
- Husain, S., Kalim, A., & Ansari, A. (2023). Sentiment analysis in social media using deep learning techniques. *International Journal of Intelligent Systems and Applications in Engineering*, 11(4), 557-568.
- Kontokosta, C. E. (2012). The price of victory: The impact of the Olympic Games on residential real estate markets. *Urban Studies*, 49(5), 961-978. <https://doi.org/10.1177/0042098011411952>
- Pierleoni, M. R., Dussauge, M., Campa, R., Regoli, A., & Anselmi, S. (2017). Assessing the Olympic Games: The economic impact and beyond. *Journal of Economic Surveys*, 31(3), 738-769. <https://doi.org/10.1111/joes.12169>
- Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M. (2021). Studying Reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media + Society*, 7(2), 1-14. <https://doi.org/10.1177/20563051211019004>
- Puschmann, C., & Powell, A. (2018). Turning words into consumer preferences: How sentiment analysis is framed in research and the news media. *Social Media + Society*, 4(3), 1-15. <https://doi.org/10.1177/2056305118797724>

Song, H., Zhang, M., & Deng, T. (2022). A systematic review of social media-based sentiment analysis: Emerging trends and challenges. *Decision Analytics Journal*, 3, 100073. <https://doi.org/10.1016/j.dajour.2022.100073>

Talha, M., Sohail, A., Tariq, R., & Ahmad, M. T. (2019). Sentiment analysis in social media and its application: Systematic literature review. *Procedia Computer Science*, 154, 76-85. <https://doi.org/10.1016/j.procs.2019.06.012>

Xu, Q. A., Chang, V., & Jayne, C. (2022). A systematic review of social media-based sentiment analysis: Emerging trends and challenges. *Decision Analytics Journal*, 3, 100073. <https://doi.org/10.1016/j.dajour.2022.100073>

Zhang, Y., Li, W., & Chen, S. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. *Artificial Intelligence Review*, 57(4), 1-45. <https://doi.org/10.1007/s10462-024-10674-8>