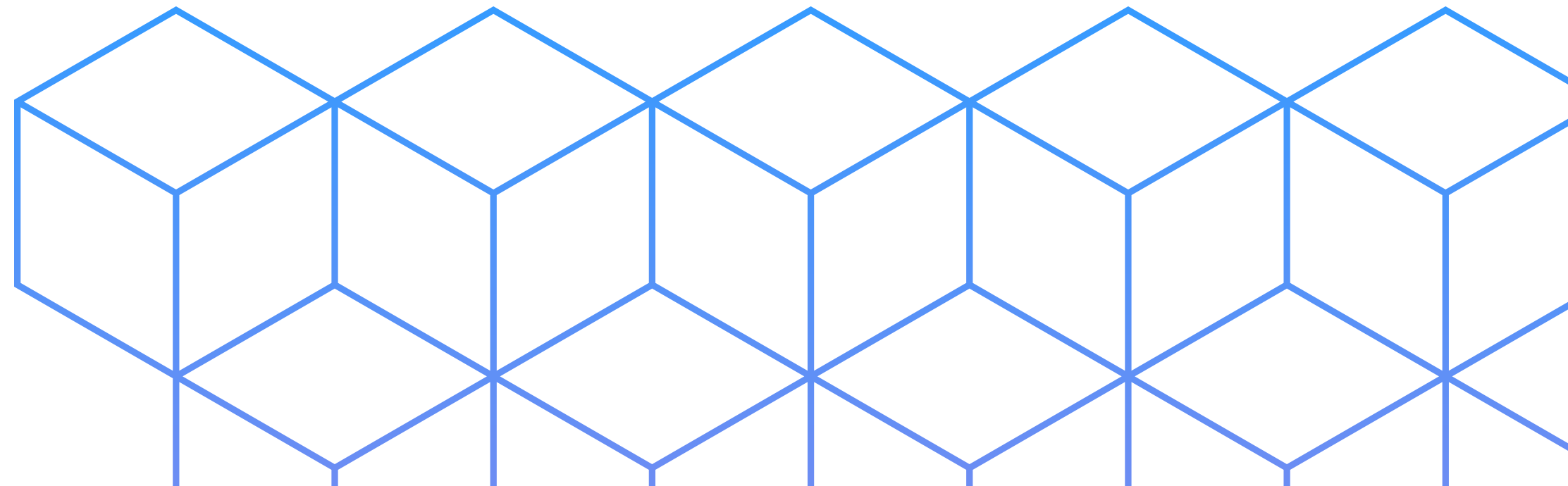


# 機器學習期末專題

## 巴黎奧運對社群情緒與經濟影響之分析

組員：張逸安 周聖詠 謝蕙宇 黃裕媿



Introduction 簡介

Related Work 相關文獻探討

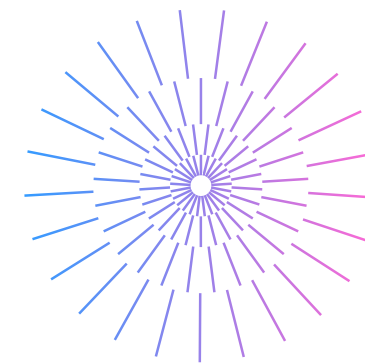
Data Description 資料說明

Methodology 研究方法

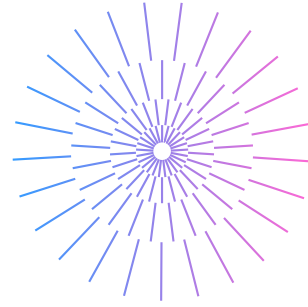
Results & Analysis 實驗結果與分析

Discussion 討論

Conclusion & Future Work 結論與未來工作



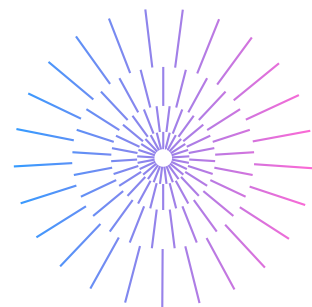
# Agenda



# Introduction

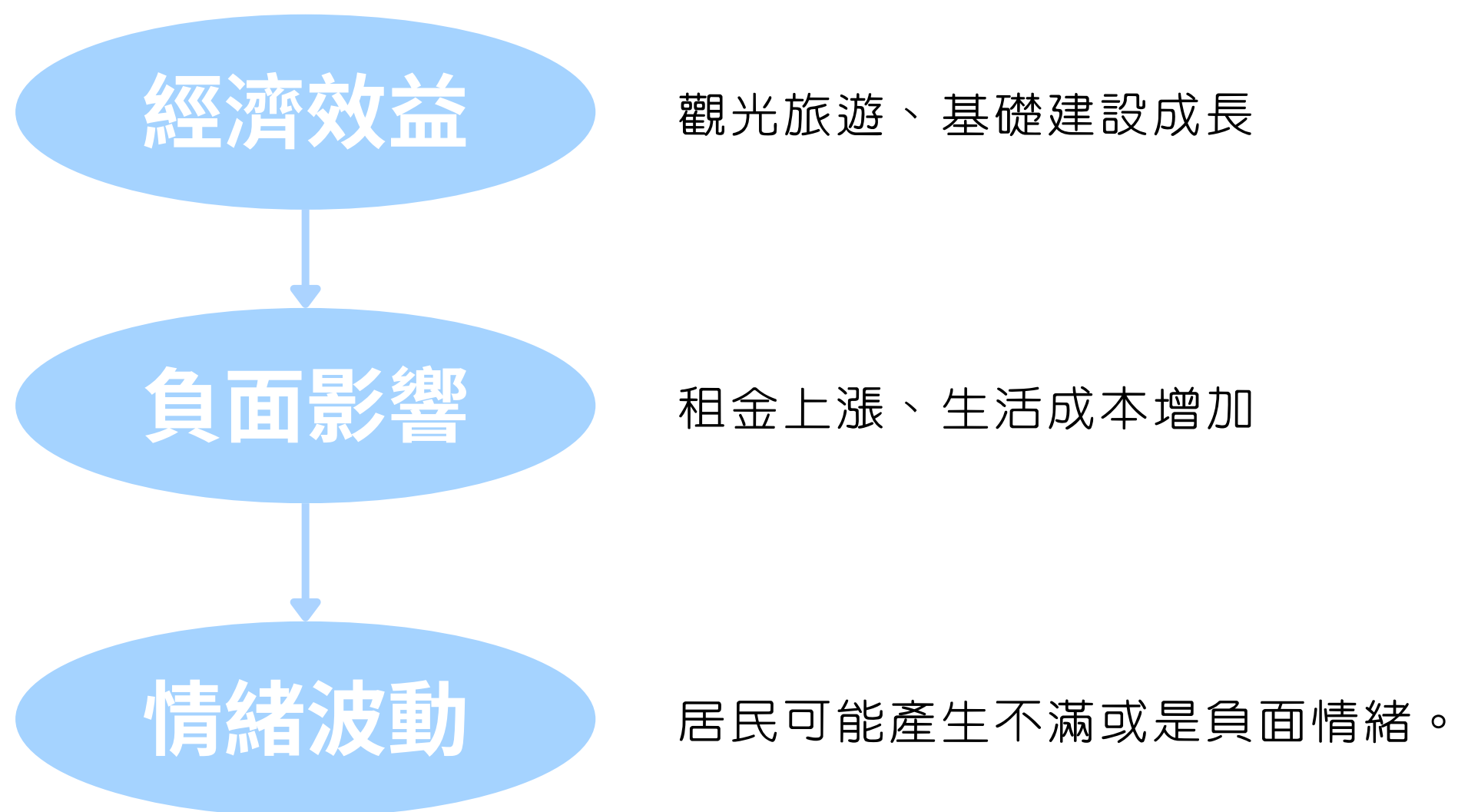
## 簡介

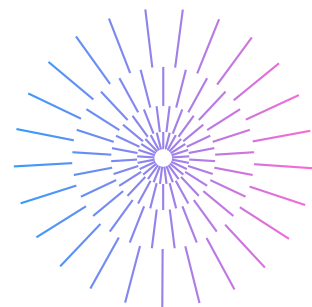
- 研究背景
- 問題陳述



# 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





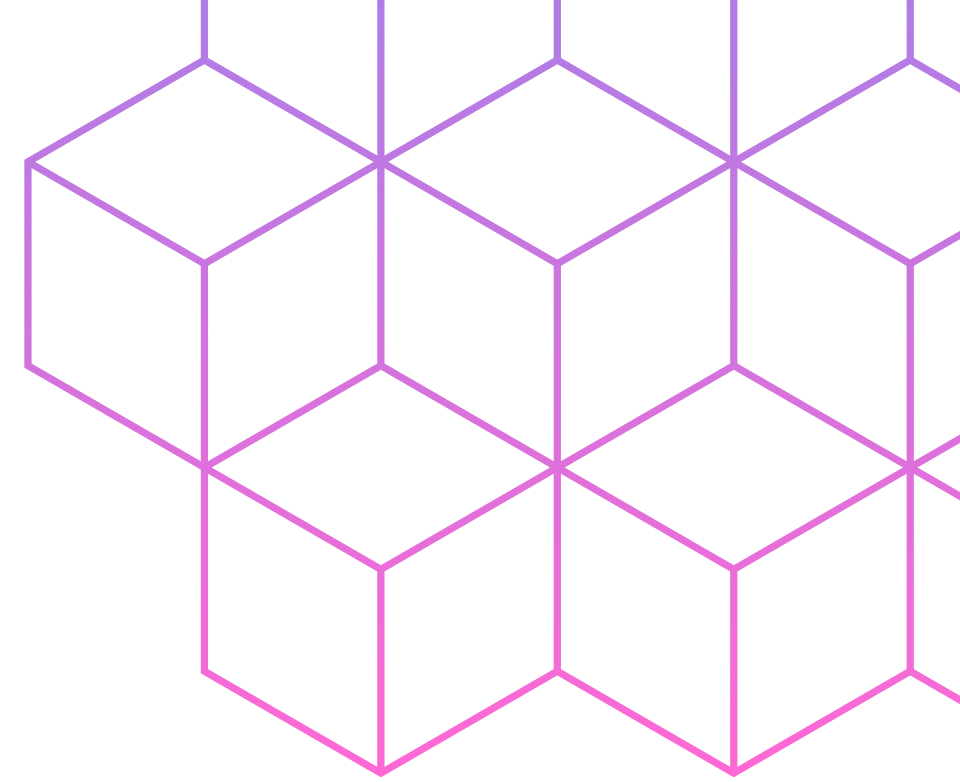
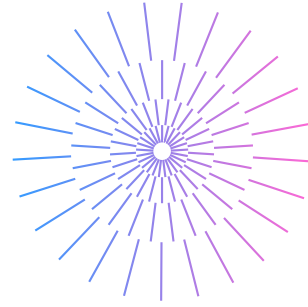
# 問題陳述

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？

預期基於研究結果，提供政策建議：

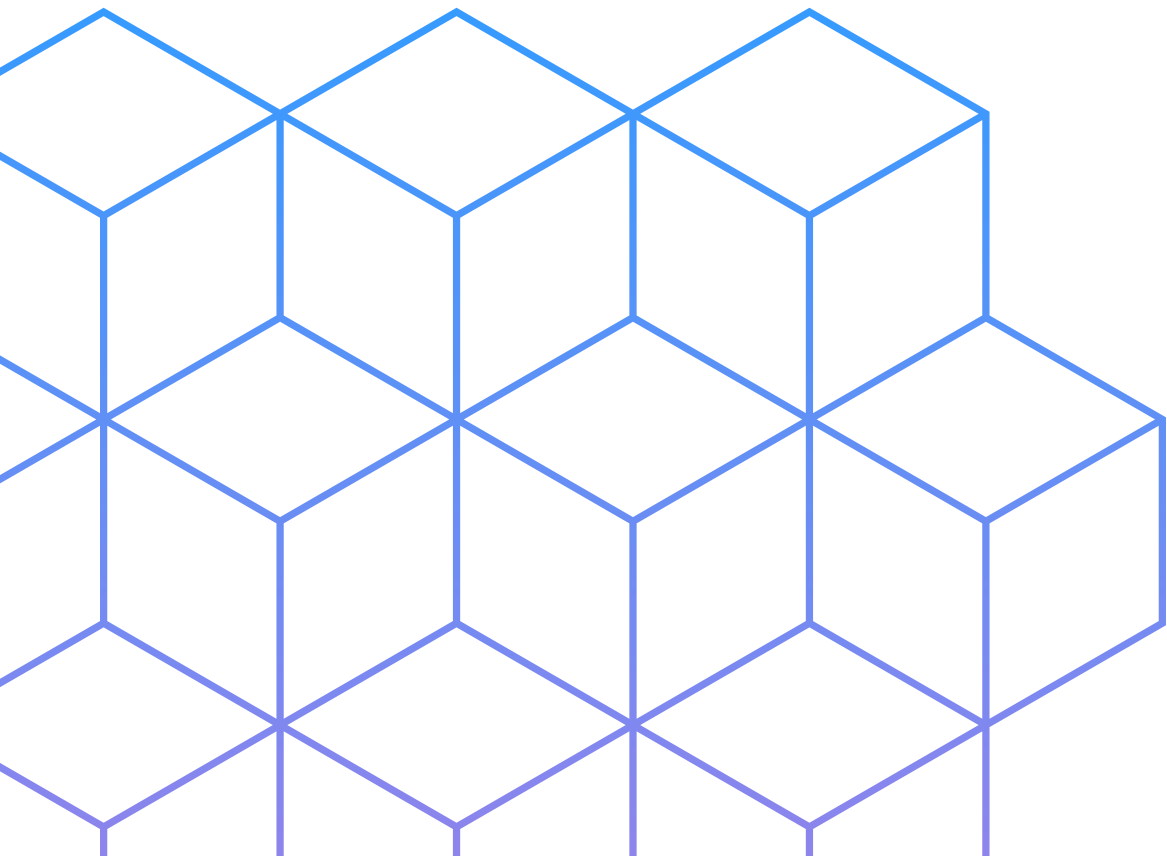
1. 改善城市居民經濟福祉之建議
2. 針對情緒負面影響提出緩解策略
3. 為未來奧運主辦城市提供經濟與情緒管理之最佳實踐

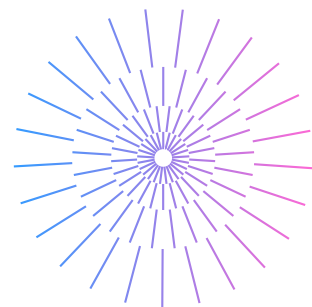


# Related Work

## 相關文獻探討

- 奧運對主辦城市的經濟影響
- 社群媒體情緒分析研究
- 經濟因素與公眾情緒之交集
- 研究缺口與貢獻





# 奧運對主辦城市 經濟影響

Baade 和 Matheson (2016) 提供奧運成本效益的全面性分析  
在大多數情況下，奧運對主辦城市而言是虧損的投資

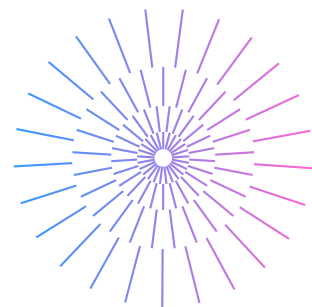
住房市場效應

2024 年巴黎奧運提供住房市場動態的案例研究  
奧運期間住宿平均每晚價格比 2023 年同期高出五  
倍，從 141 歐元上升至 706 歐元

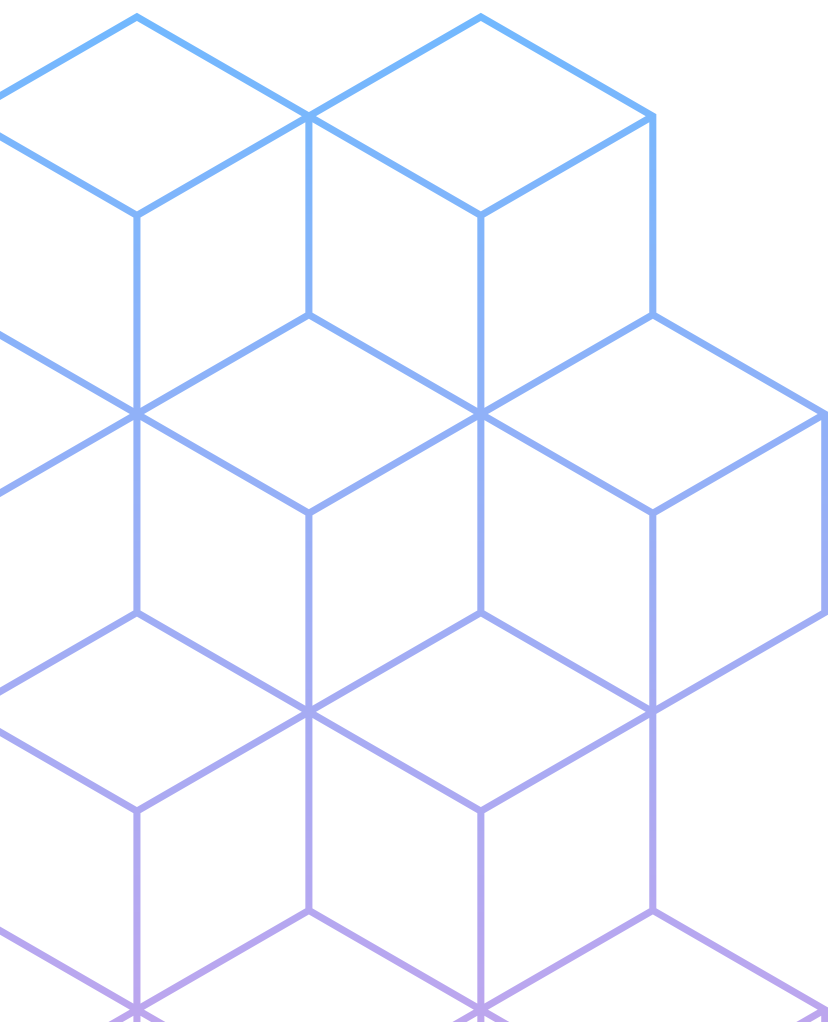
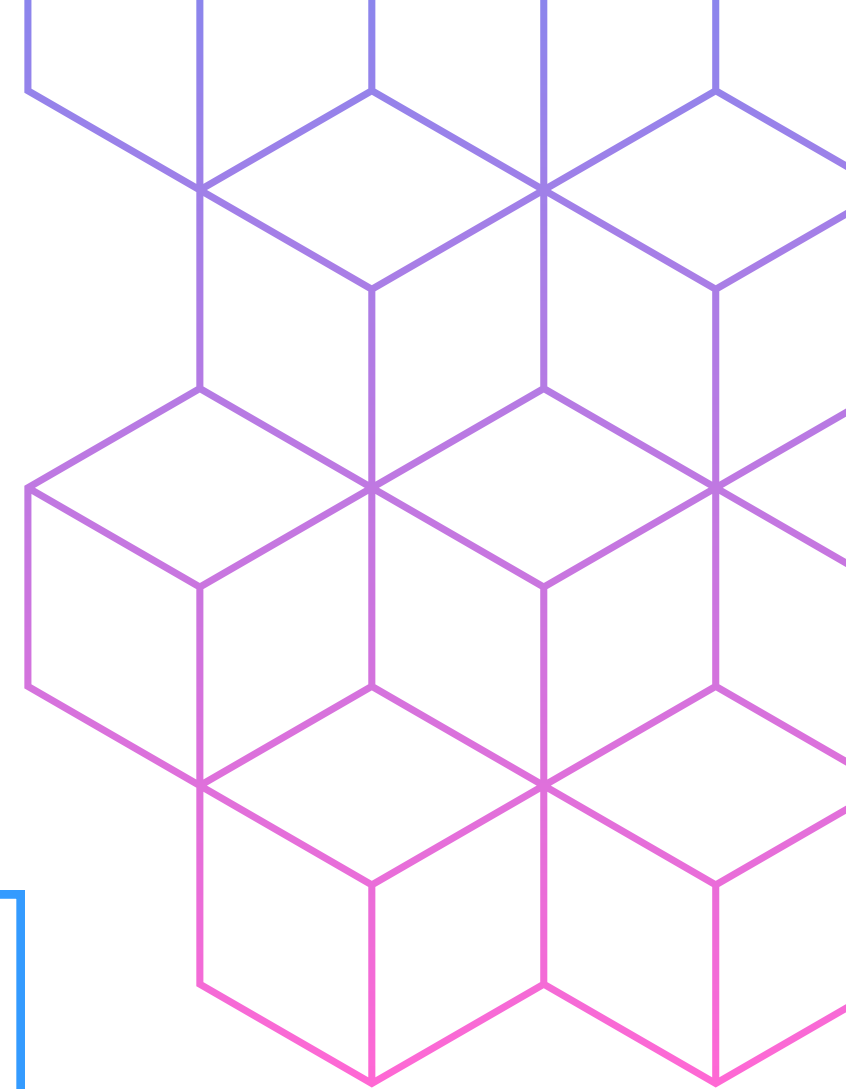
就業影響效應

- 2002 年鹽湖城奧運，失業統計數據未顯著改善
- 奧運影響結果好壞參半，就業影響未達預期  
(Billings & Holladay, 2012; Maennig & Richter,  
2012; Langer et al., 2018)





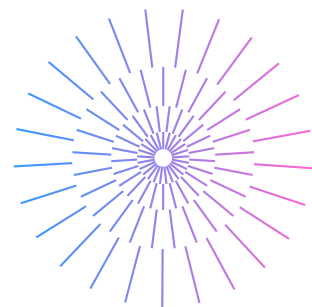
# 社群媒體 情緒分析研究



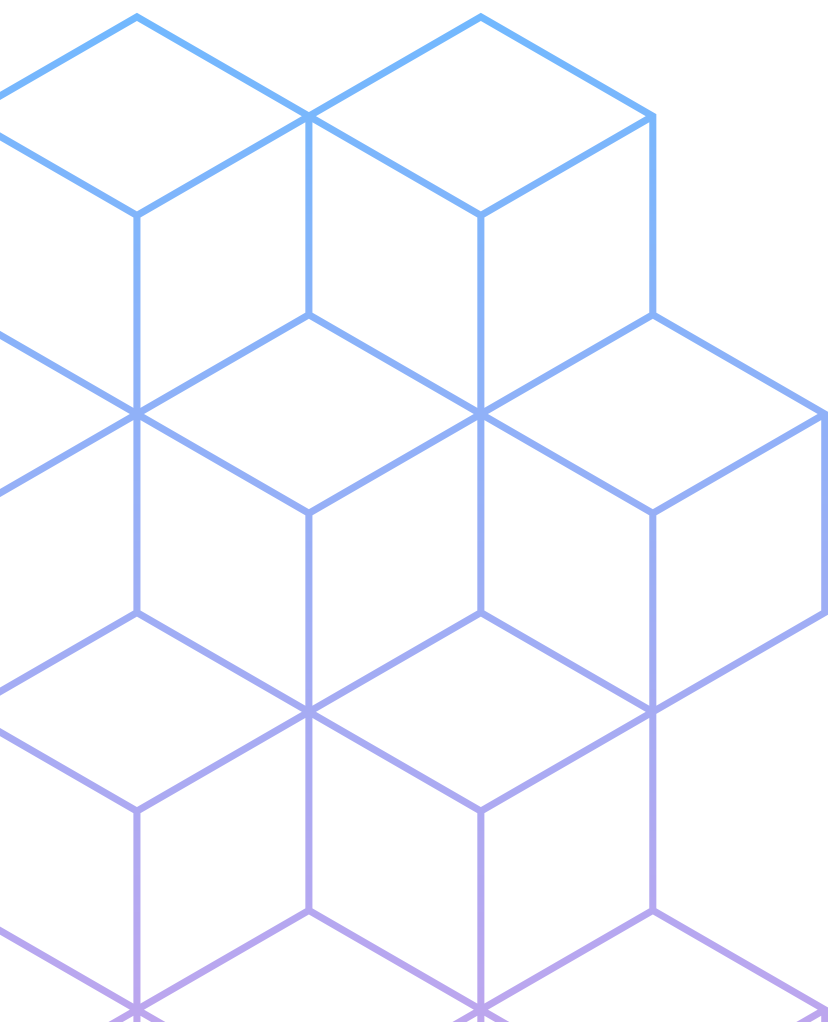
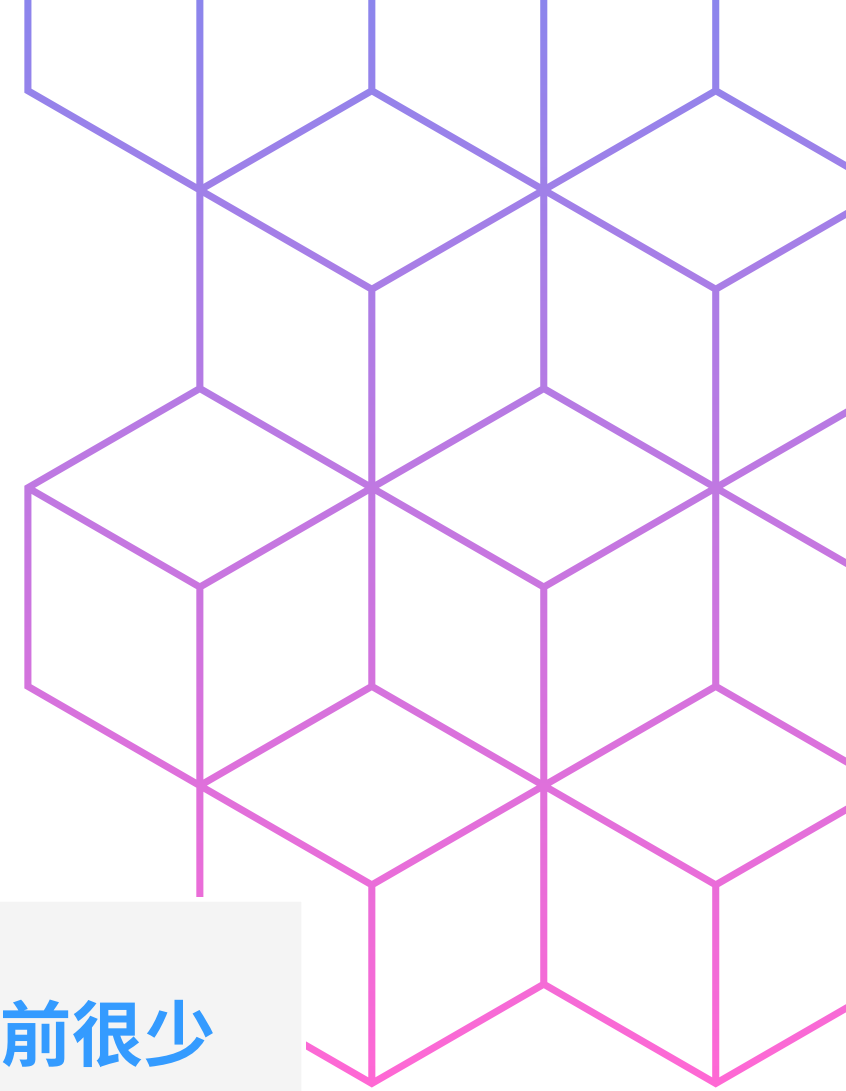
情緒分析領域已從基本的極性檢測，發展為結合時間動態、網路關係和情緒強度檢測的複雜多模態分析。

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)。



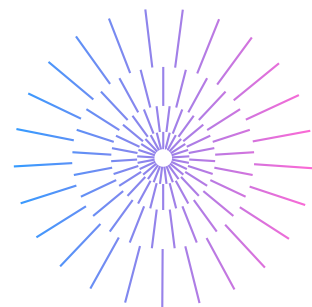


# 經濟因素與 公眾情緒之交集



雖然奧運經濟影響和社群媒體情緒分析分別有大量研究，目前很少研究試圖在大型活動期間建立經濟變數與公眾情緒的聯繫：

- 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)
- 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)



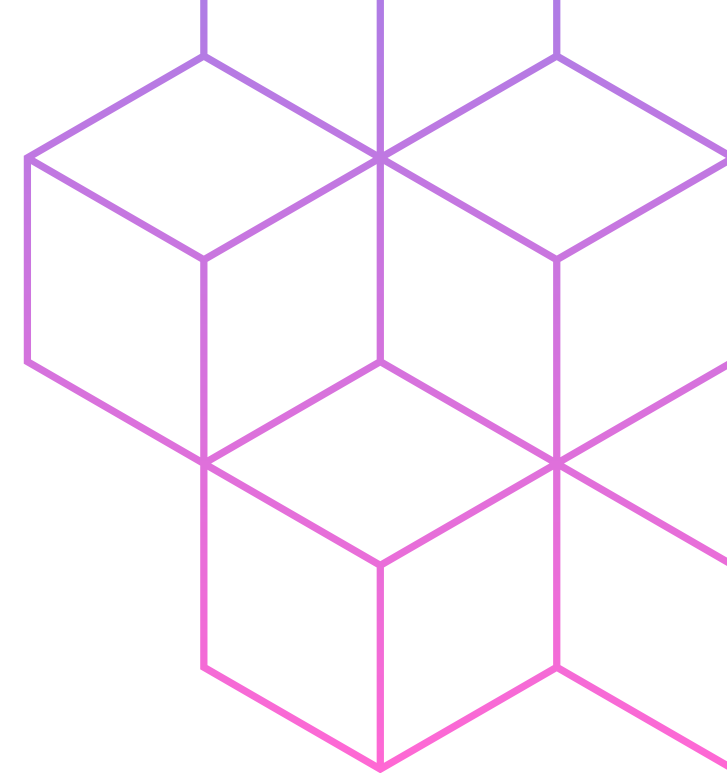
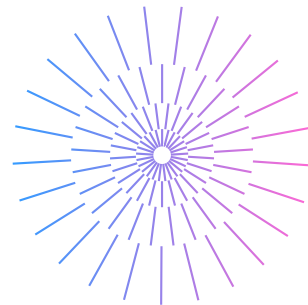
# 研究缺口與貢獻

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

本研究主要貢獻包含

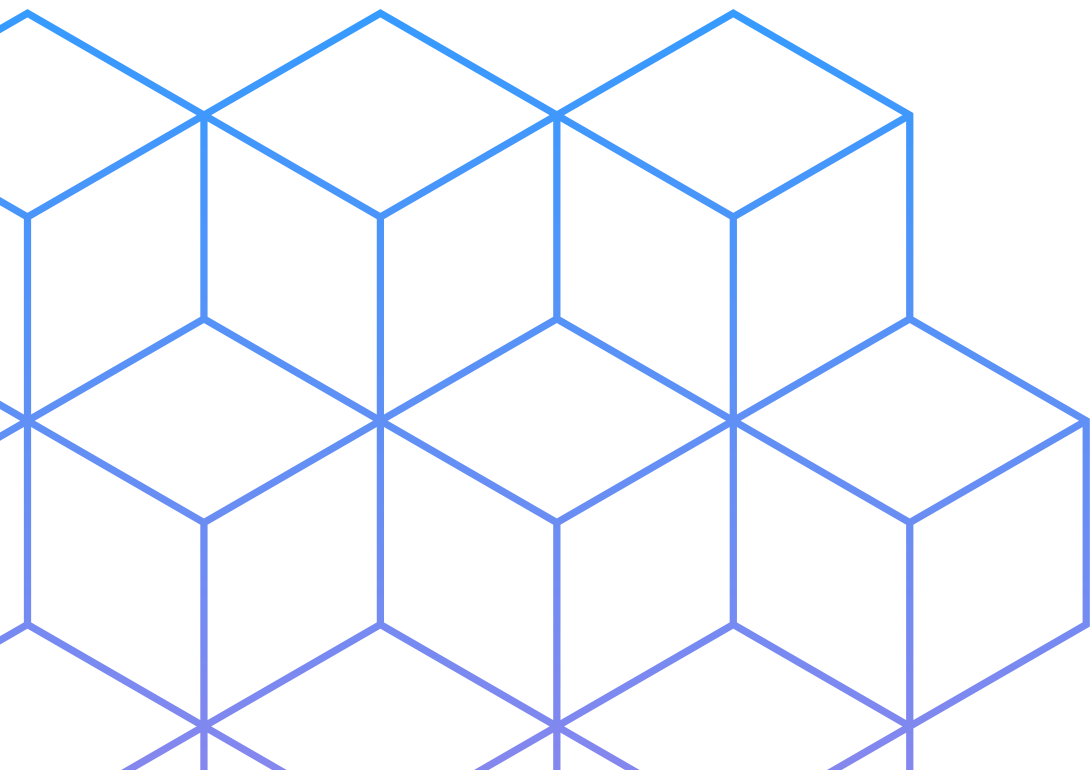
- (1) 在奧運影響評估中開創性整合經濟（Airbnb 價格）和情緒（Reddit 評論）資料
- (2) 應用機器學習技術從經濟指標預測情緒
- (3) 提供關於大型活動期間短期住房市場波動與公眾情緒之間關係的實證證據

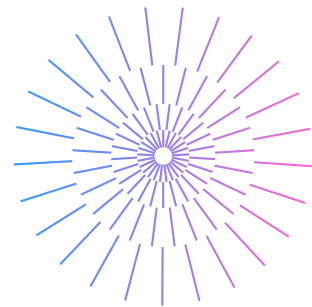


# Data Description

## 資料說明

- 情緒資料蒐集與處理
- 經濟數據處理與分析





# 情緒資料蒐集

## Facebook

初期嘗試以 FB 為主要平台  
進行情緒資料收集

問題：

- 爬蟲工具與 API 常遭封鎖
- 教學資料過時，實作困難

## YouTube

利用關鍵字搜尋影片  
收集下方留言進行情緒分析

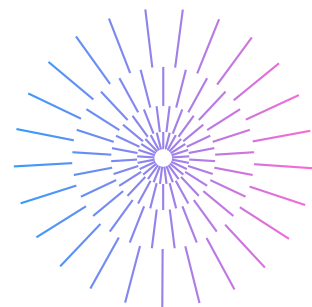
問題：

嘗試多種模型，但  $R^2$  接近 0  
模型幾乎無預測力

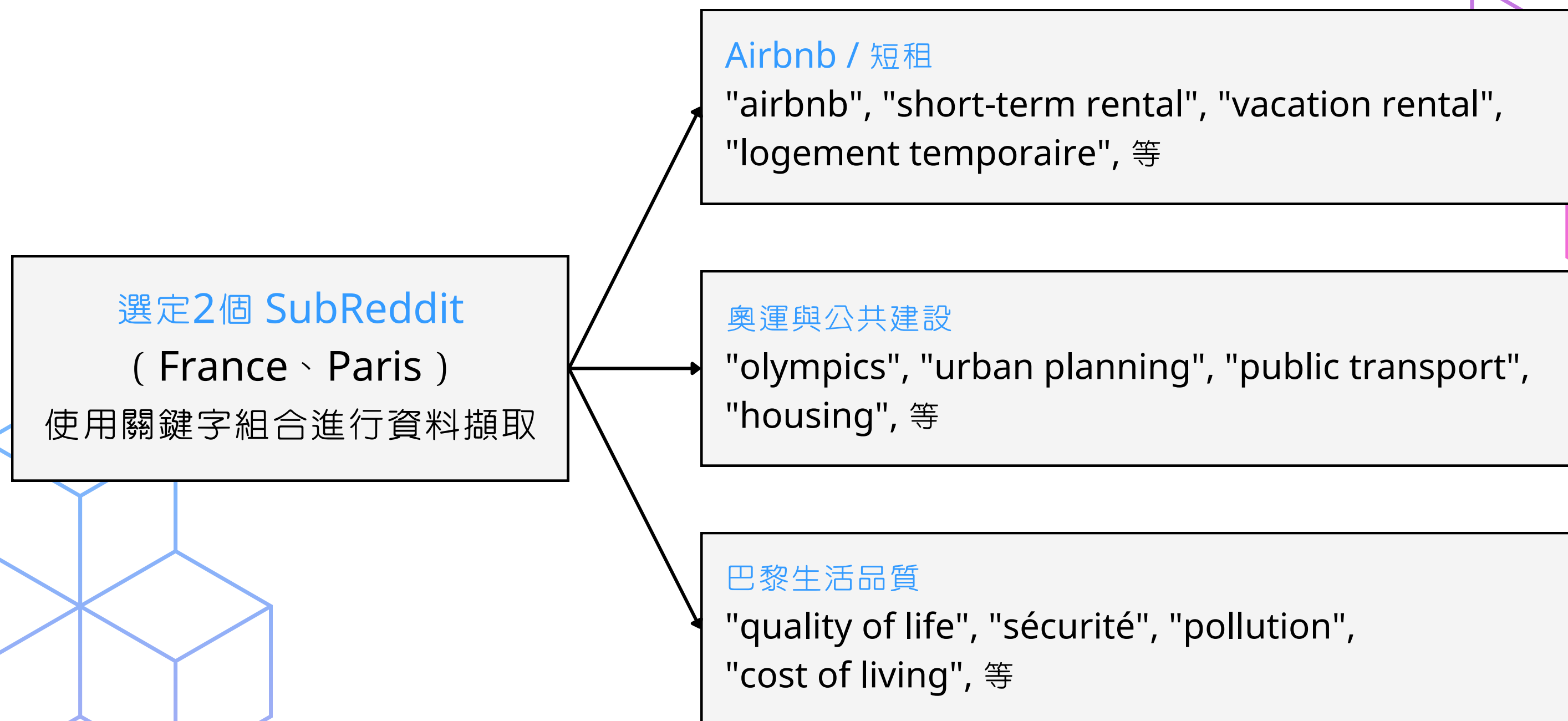
## Reddit

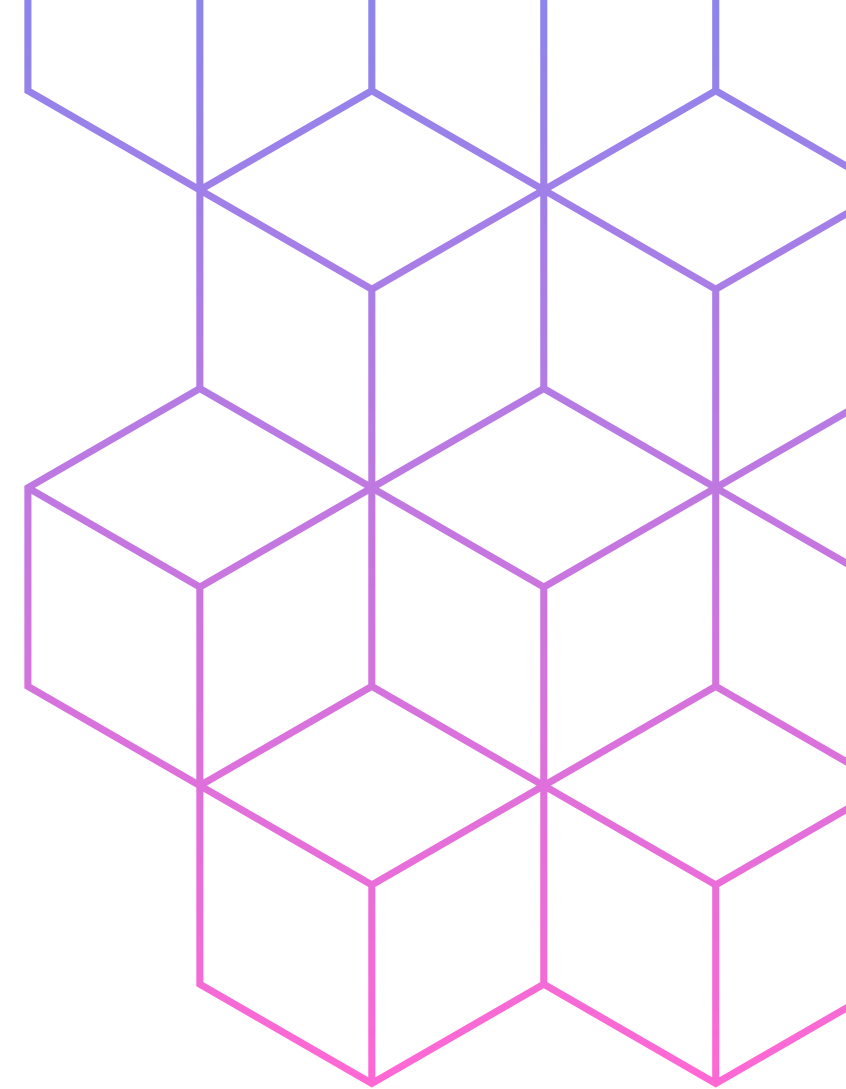
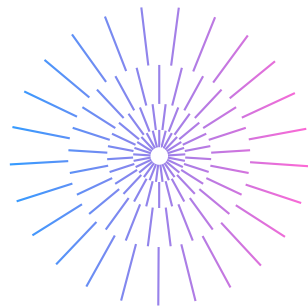
選擇理由：

- 爬蟲友善、資料量大
- 年輕族群集中
- 對房價與城市生活議題  
參與度高



# 情緒資料處理





## reddit\_sensor\_priority.ipynb

技術流程：

Google Translate API 進行非英文留言的翻譯

TextBlob 進行英文留言的情緒極性分析

## translated\_sentiment\_full.csv

爬蟲留言經翻譯與情緒分析後匯整而成的完整結果

## airbnb\_paris\_translated\_sentiment.csv

篩選出Airbnb/房價相關之留言子集

保留情緒分析結果

用途：觀察居住壓力與社會情緒變化的關聯

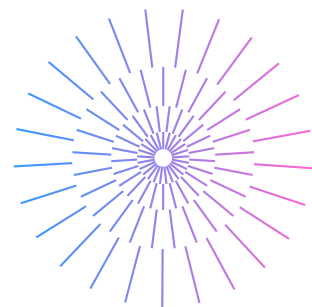
## olympics\_topic\_translated\_sentiment.csv

篩選出與olympics相關之留言子集

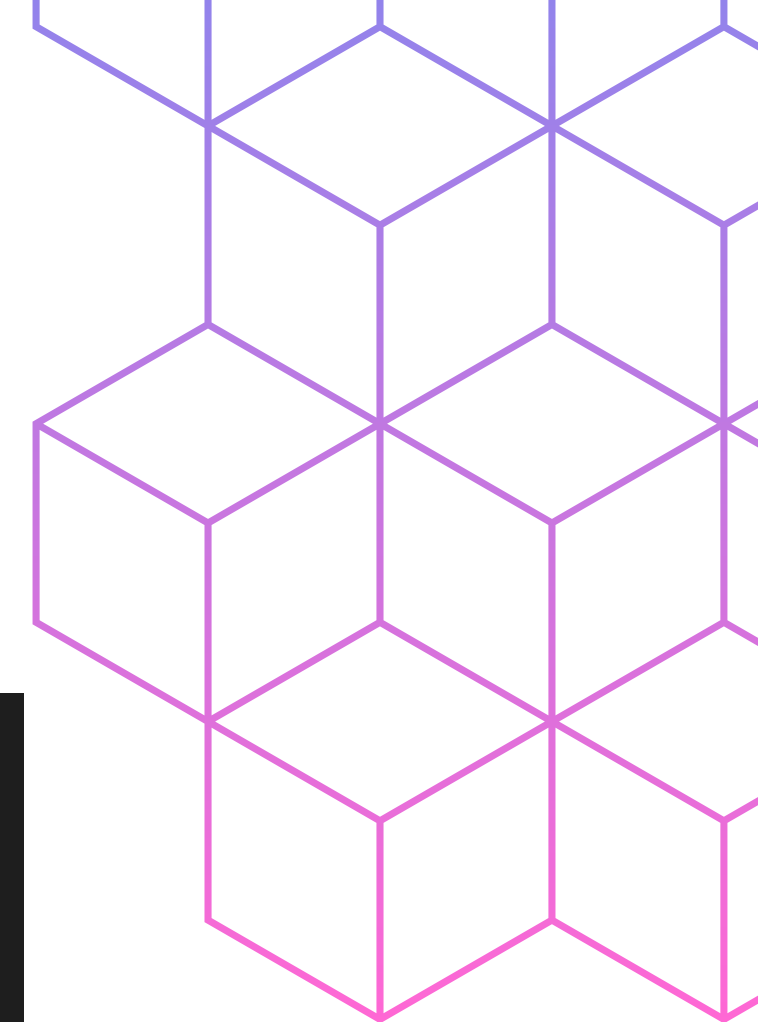
保留情緒分析結果

用途：分析巴黎奧運對社會情緒的影響。



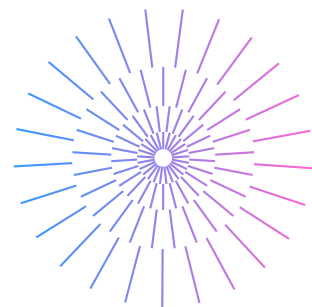


# Airbnb數據處理



	listing_id	date	available	price	adjusted_price	minimum_nights	maximum_nights
0	3109	2023-09-05	f	\$110.00	\$110.00	2.0	30.0
1	3109	2023-09-06	f	\$110.00	\$110.00	2.0	30.0
2	3109	2023-09-07	f	\$110.00	\$110.00	2.0	30.0
3	3109	2023-09-08	t	\$110.00	\$110.00	2.0	30.0
4	3109	2023-09-09	t	\$110.00	\$110.00	2.0	30.0
5	3109	2023-09-10	t	\$110.00	\$110.00	2.0	30.0
6	3109	2023-09-11	f	\$110.00	\$110.00	2.0	30.0
7	3109	2023-09-12	f	\$110.00	\$110.00	2.0	30.0
8	3109	2023-09-13	f	\$110.00	\$110.00	2.0	30.0
9	3109	2023-09-14	f	\$110.00	\$110.00	2.0	30.0
10	3109	2023-09-15	t	\$110.00	\$110.00	2.0	30.0





# Airbnb數據處理

## Airbnb price change trend analysis

每日平均房價變化



- 圖中2024/7/1為第一次驟升點  
以此作為奧運前後的分界
- 整理出以下，作為價格全面特徵：

平均價格變化

avg\_price\_change

價格變化波動(標準差)

std\_price\_change

平均價格

avg\_price

價格中位數

median\_price

價格範圍(最大值-最小值)

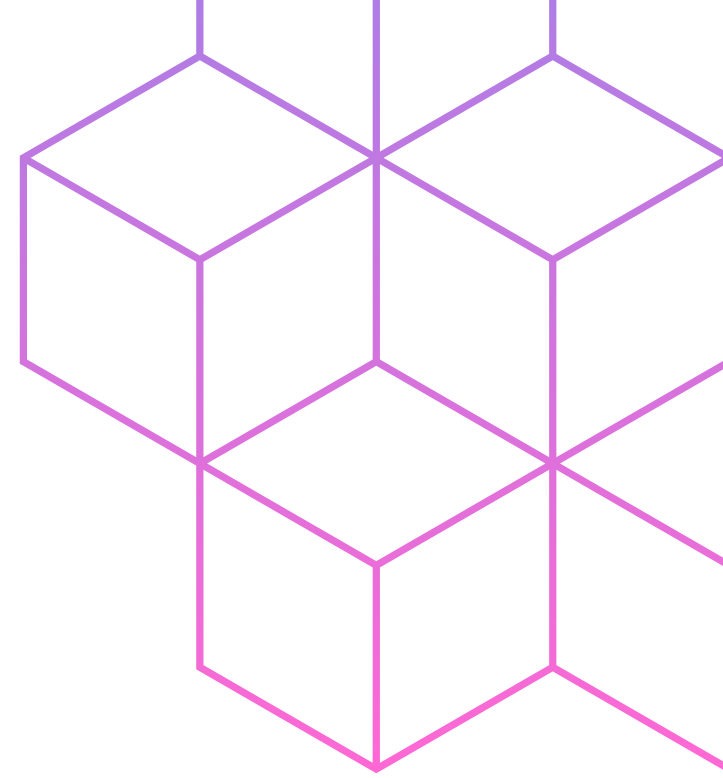
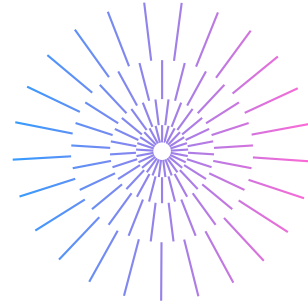
price\_range

平均價格對數

log\_avg\_price

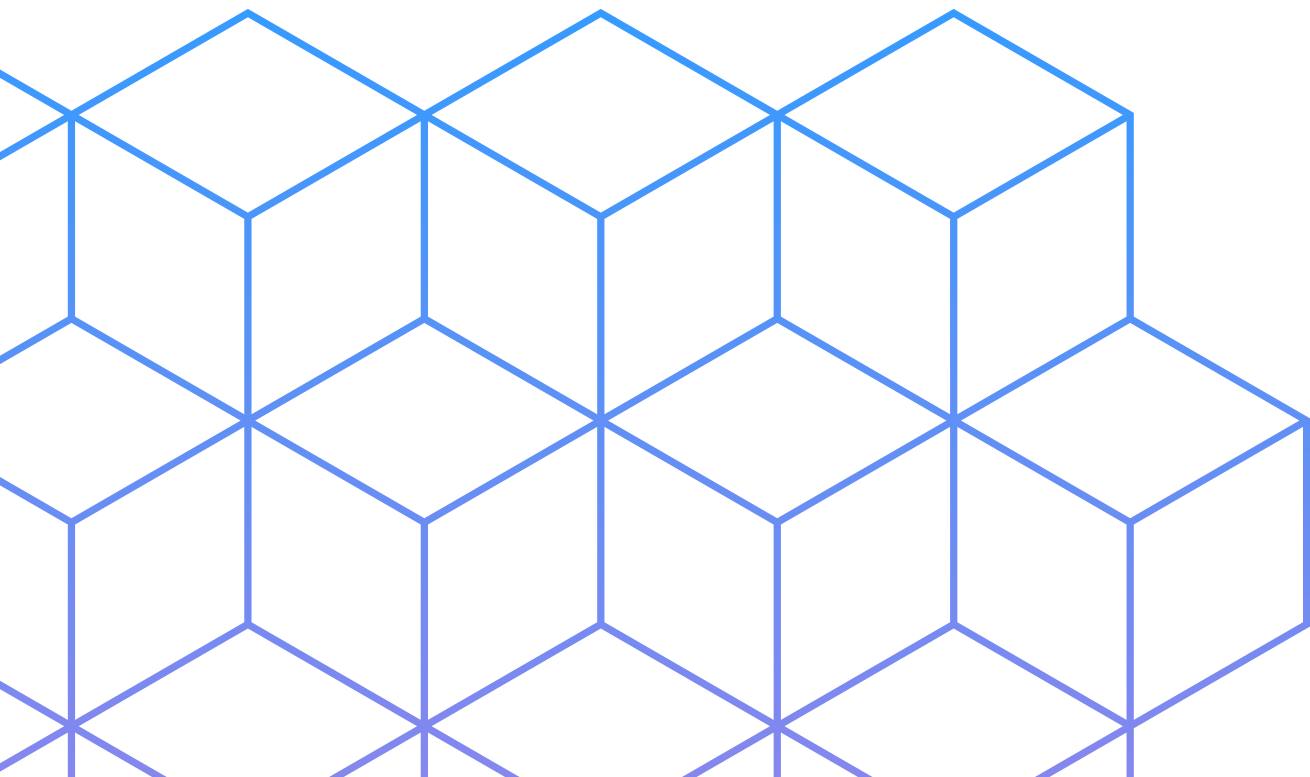
相對價格變化百分比

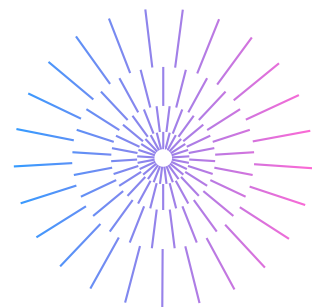
relative\_price\_change



# Methodology 研究方法

- 模型選擇
- 實作流程
- 特徵工程





# 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：

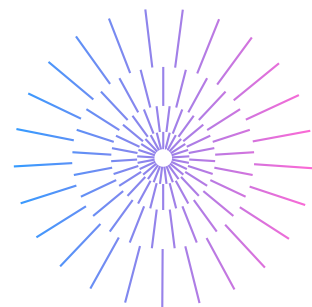
Linear Regression、Ridge、Lasso、ElasticNet

## 樹模型：

Decision Tree、Random Forest、Gradient Boosting、AdaBoost

## 其他模型：

KNN、SVM、XGBoost、LightGBM



# 實作流程

自變數

8個價格特徵

目標變數

留言情緒分數

執行重要性分析，解釋預測情緒變化的重要價格特徵

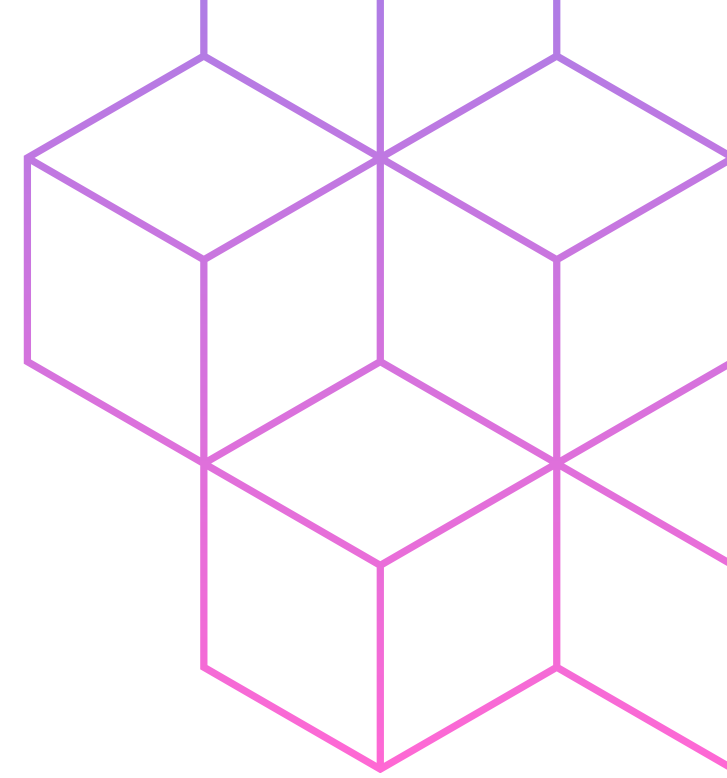
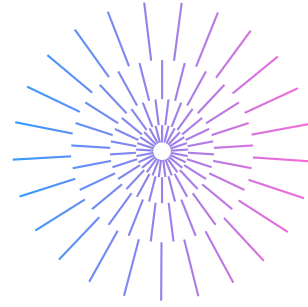
結果：

將未經過處理的資料進行訓練  
判定係數 $R^2$ 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

特徵工程

Airbnb 相關留言

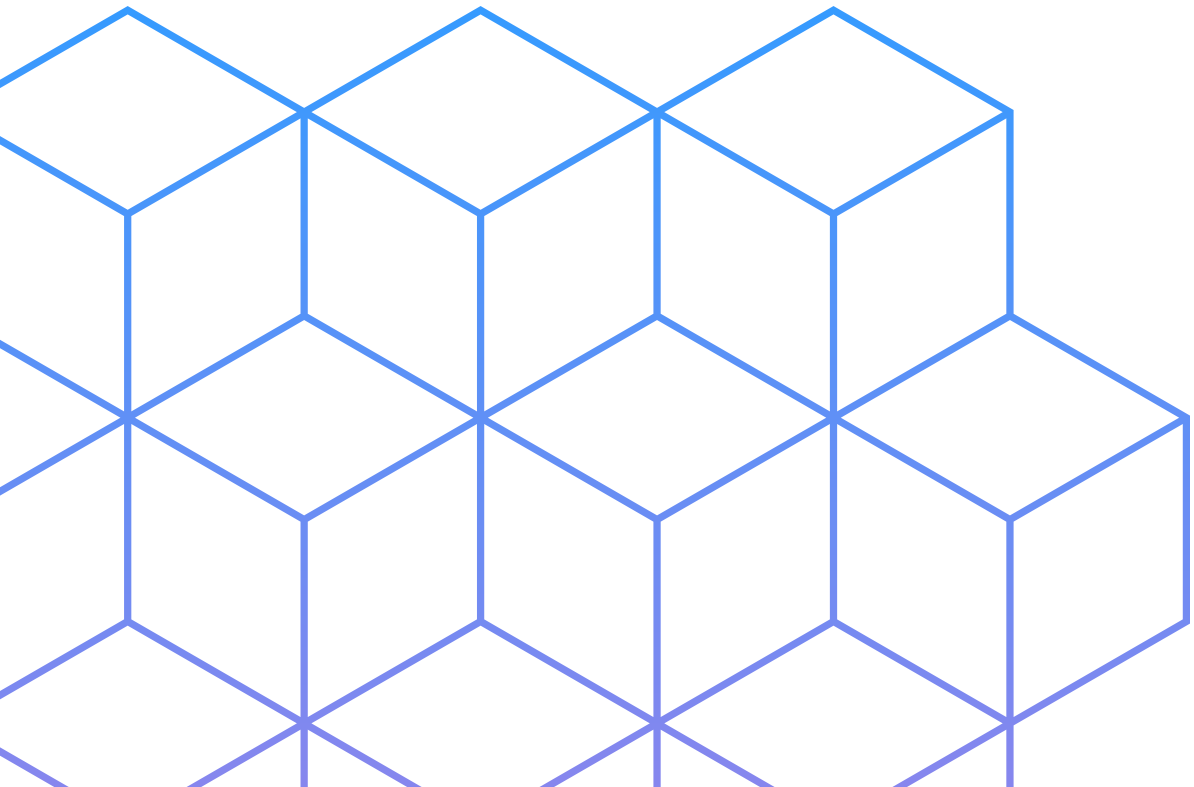
Olympics 相關留言

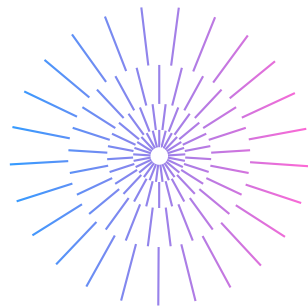


# Results & Analysis

## 實驗結果與分析

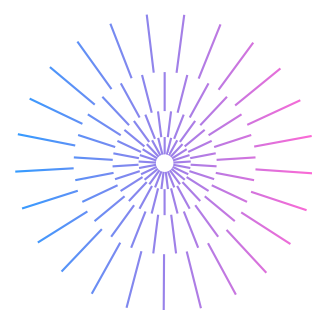
- 模型整體表現
- 視覺化分析





# 模型整體表現

Model	特徵工程前R <sup>2</sup>	Airbnb相關R <sup>2</sup>	Olympic相關R <sup>2</sup>
Linear Regression	0.0006	0.0021	-0.0003
Ridge Regression	0.0007	0.0018	-0.0002
Lasso Regression	-0.0002	0	-0.0002
ElasticNet	-0.0002	0	-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	0.002	-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

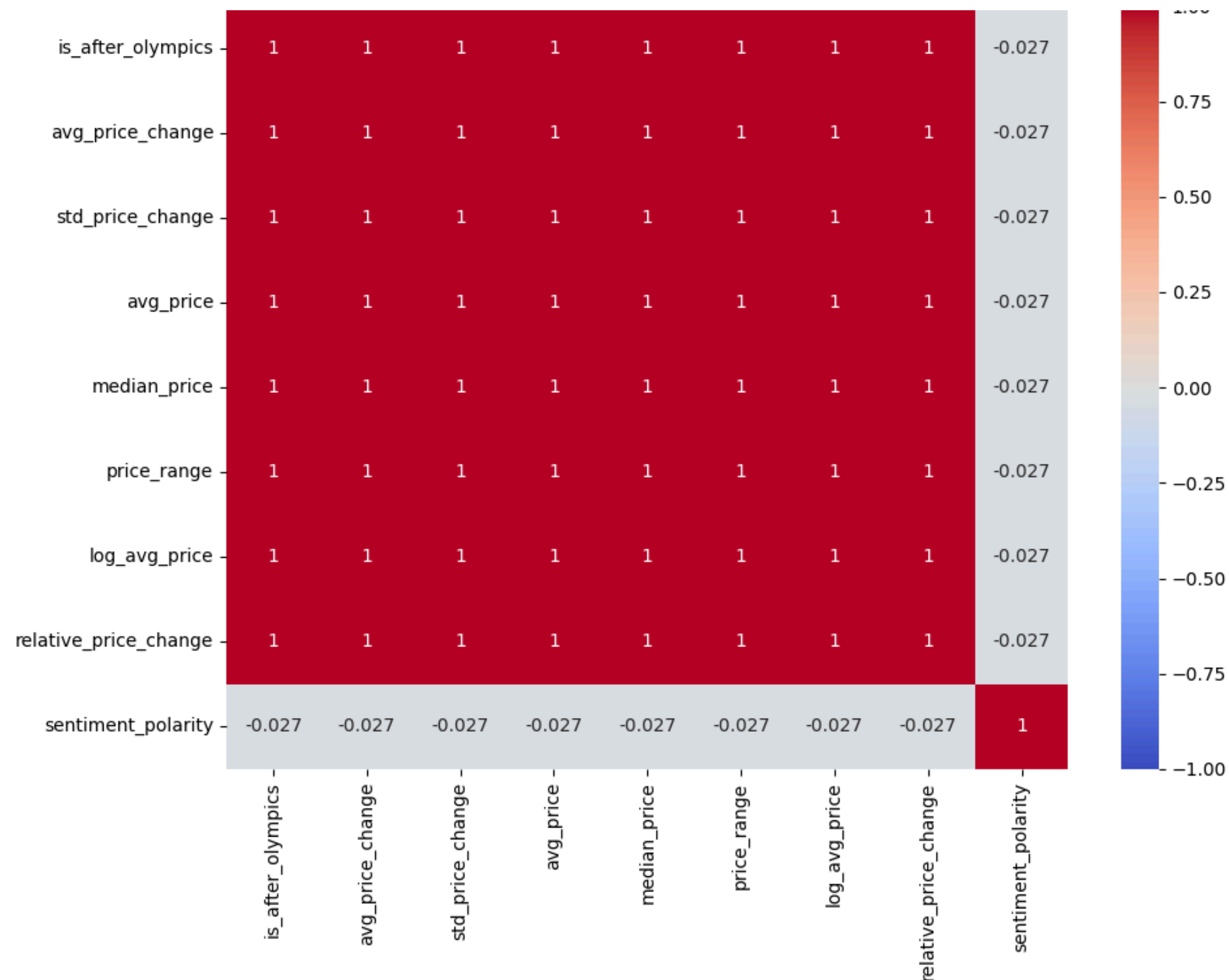


# 視覺化分析

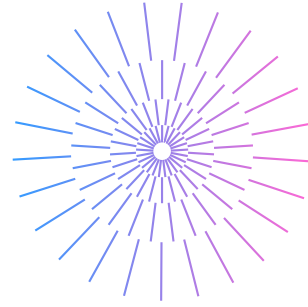
Airbnb 價格特徵與  
情緒分數之間的相關性熱圖



兩者之間沒有關係

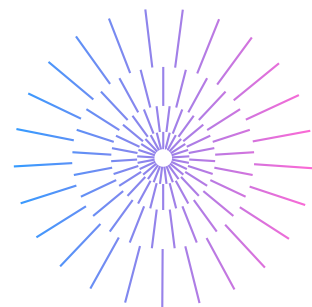






# Discussion 討論

- 結果意義與反思
- 資料與模型的限制



# 結果意義與反思

社群的情緒與房價變化之間缺乏關聯，相關係數皆接近 0，迴歸模型預測力有限  
這可能反映出以下問題：

01

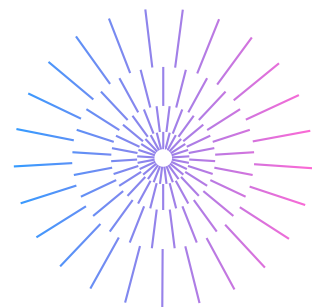
兩者之間本就沒有明確可預測的關係

02

變數間的關係並非線性

03

情緒表現不是單靠價格變化就能解釋  
更多受到當下語氣、自身想法、價值觀所影響



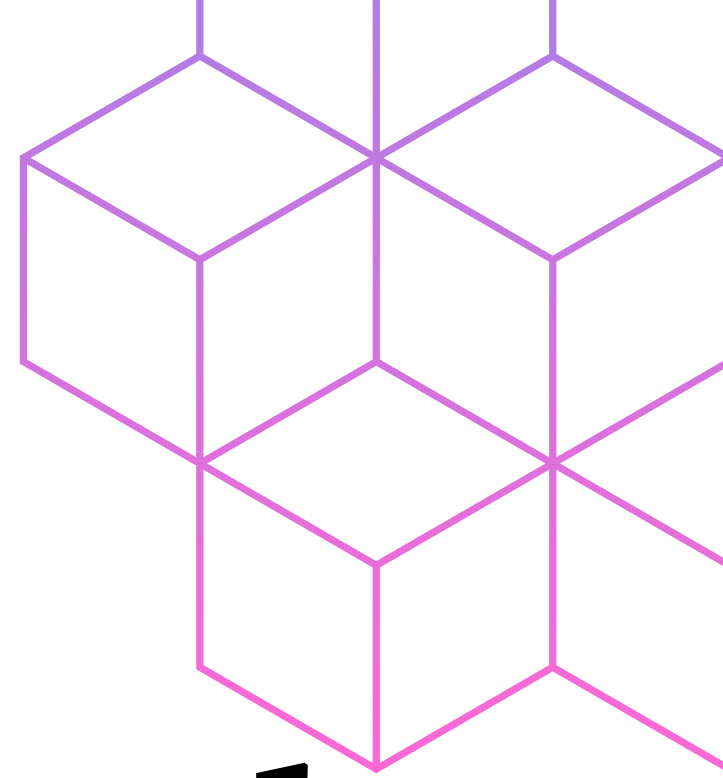
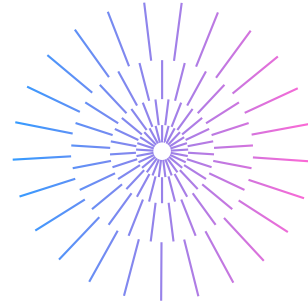
# 資料與模型的限制

資料來源與樣本限制

情緒分析工具限制

特徵工程的不足

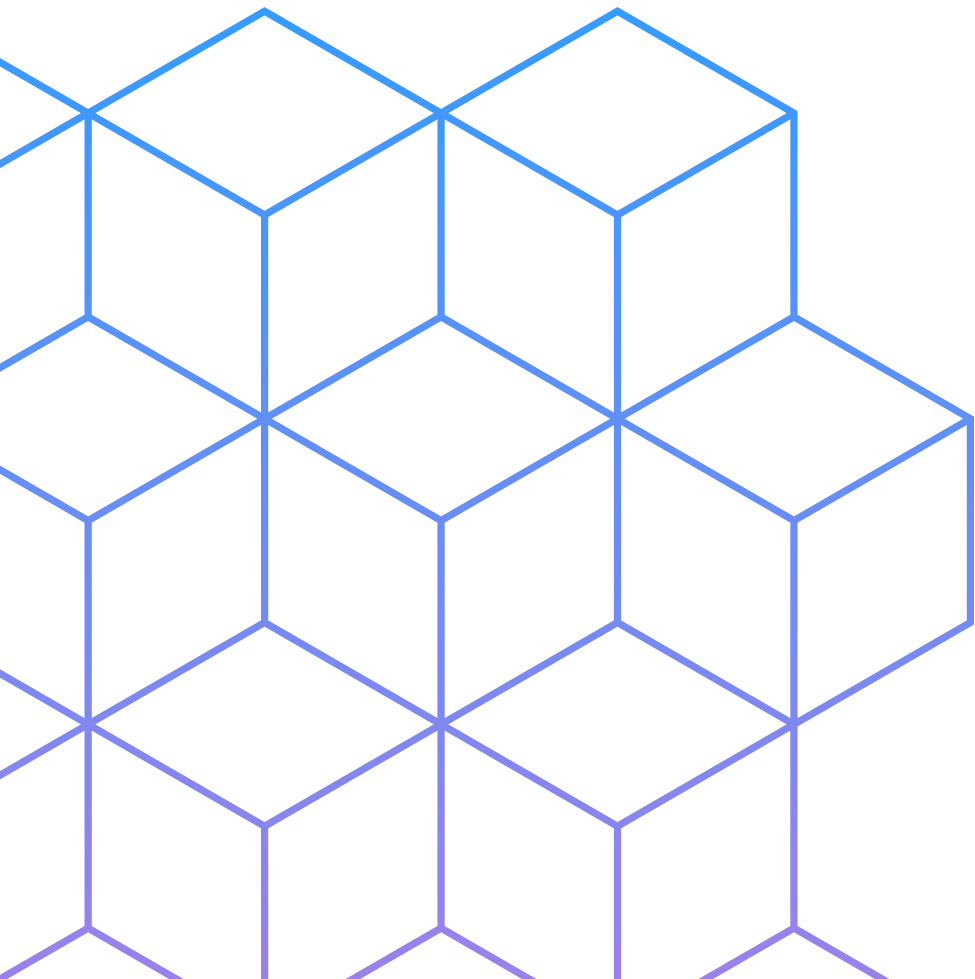
建模方向與挑戰

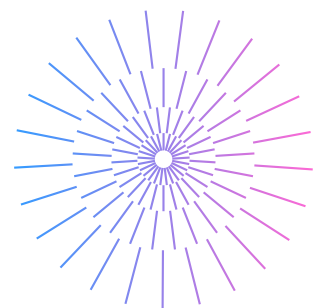


# Conclusion & Future Work

## 結論與未來工作

- 研究總結
- 未來工作與改進方向





# 研究總結

## 實驗結果

模型預測力極低， $R^2$  幾乎為 0

嘗試多種回歸模型 + 特徵工程，效果依然有限

## 發現與挑戰

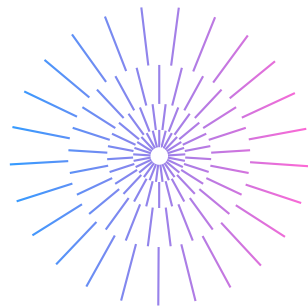
情緒為複雜且主觀現象，難以單靠價格數據建模

科學在處理主觀情緒現象時仍會面臨挑戰

## 實作啟示

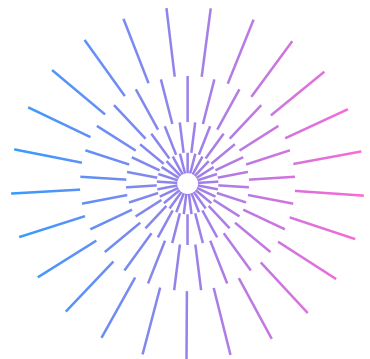
特徵工程嘗試具啟發性

加入語意、事件、互動等特徵，增加預測潛力



# 未來工作與改進方向

策略	方法	預期成果
情緒從分數改為分類	連續的分數 轉換成三分類：正/中/負	比回歸模型更能學習出模式
使用進階情緒分析模型	VADER、BERT、RoBERTa	更好地處理 非正式語言及多語背景
補強特徵設計	增加主題分類、留言熱度、 發文時間、地理位置等特徵	助於模型捕捉到關係



# 感謝聆聽

組員：張逸安 周聖詠 謝蕙宇 黃裕媿

