

線上訂房平台調價策略之研究

作者：鄭恆安

指導教授：國立臺灣大學 資訊工程系 廖世偉教授

助教：謝銘鋒博士生、廖羿瑋碩士生

摘要

有次在訂飯店時我發現，同一間飯店、房間，在不同時間訂到的價錢可能都不一樣，這引起我對飯店調價策略的好奇。我利用訂房網站的數據建立自己的資料庫，並嘗試用機器學習來建立房價預測模型，來模擬飯店的調價策略。在訓練模型的過程中，發現大家常用的資料歸一化訓練的模型結果非常不理想，所以我依資料的特性提出新的歸一化方法，大大提升模型預測結果的準確性。目前針對給定的入住日，利用過去歷史 n 日訂房的房價來預測隔日訂房房價的平均絕對值誤差已可降到1%。研究結果可以提供飯店業者動態房價調整建議，讓業者根據競爭對手的調價策略動態調價，提升飯店的訂房量與營收，或許還可以作為消費者訂房前的參考。

研究動機

- 同一個房間，在不同時間訂房的價格並不一樣。
- 如何訂到最低的房價呢？
- 實作飯店調價預測模型以預測房價。

研究目標

1

找出影響飯店調價策略之變因。

2

建構飯店調價預測模型。

名詞定義

- 間隔天數：訂房日與入住日的天數差
- MAPE：絕對值平均百分誤差，越小代表誤差越少
- 歸一化：根據資料的最大最小值，將所有數值縮放到0和1之間
- 歷史資料天數 n ：調價預測模型所需要的輸入資料長度。
- 訓練集：用來訓練模型
- 測試集：用來觀察模型表現

研究方法

建立房價
資料庫

視覺化
分析

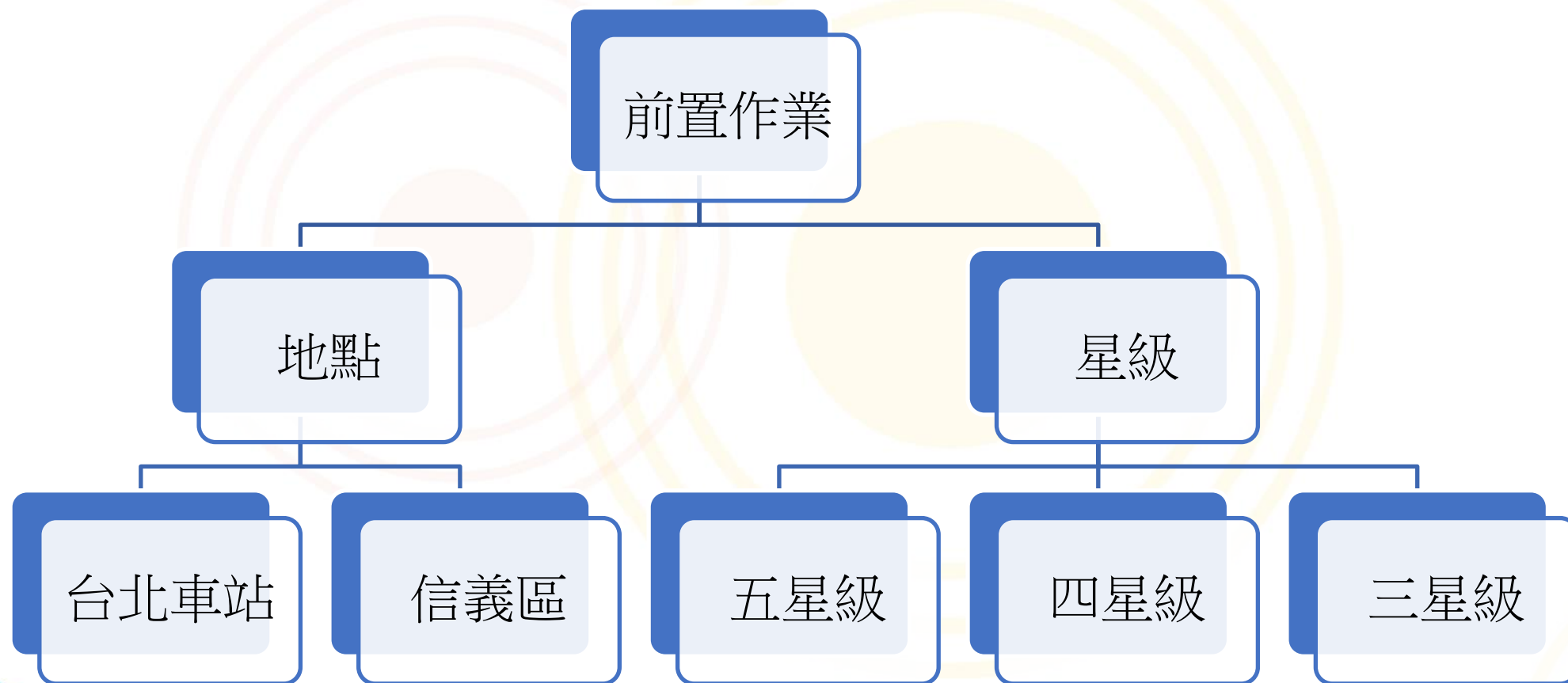
訓練模型

預測結果

The background features a light cream color with abstract geometric patterns. In the corners, there are clusters of parallel lines in yellow, red, and blue. In the center, there are two sets of concentric circles: a smaller one in light blue and a larger one in light yellow. The text '建立房價資料庫' is centered in a bold, dark blue font.

建立房價資料庫

建立房價資料庫(選定研究對象)



建立房價資料庫(爬蟲)

建立資料庫

- 透過爬蟲將所有飯店的房型存檔
- 建立每間飯店的資料庫

蒐集每日房價

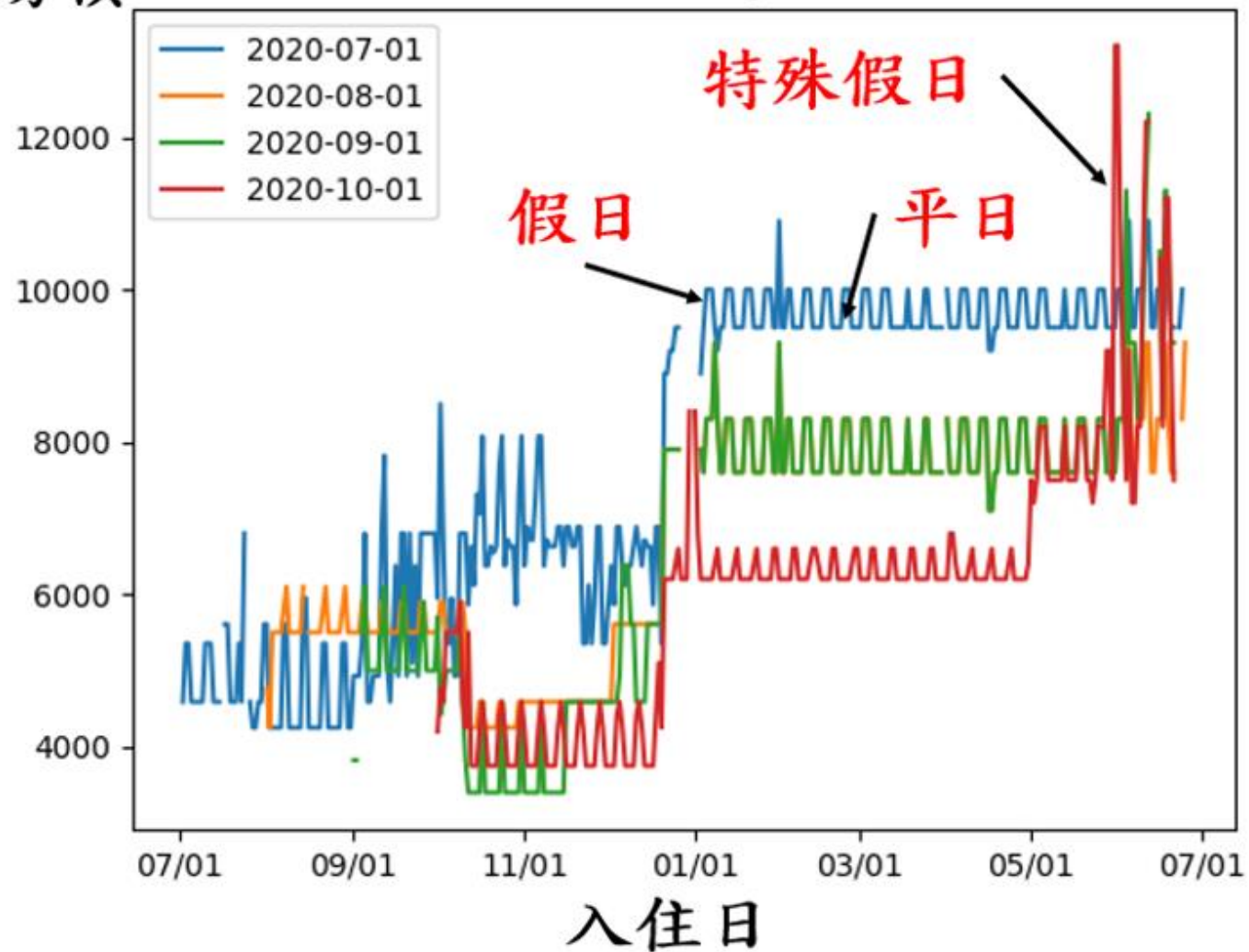
- 改變不同入住日、飯店進行搜尋
- 將相對應飯店房型的房價存入資料庫中

視覺化分析

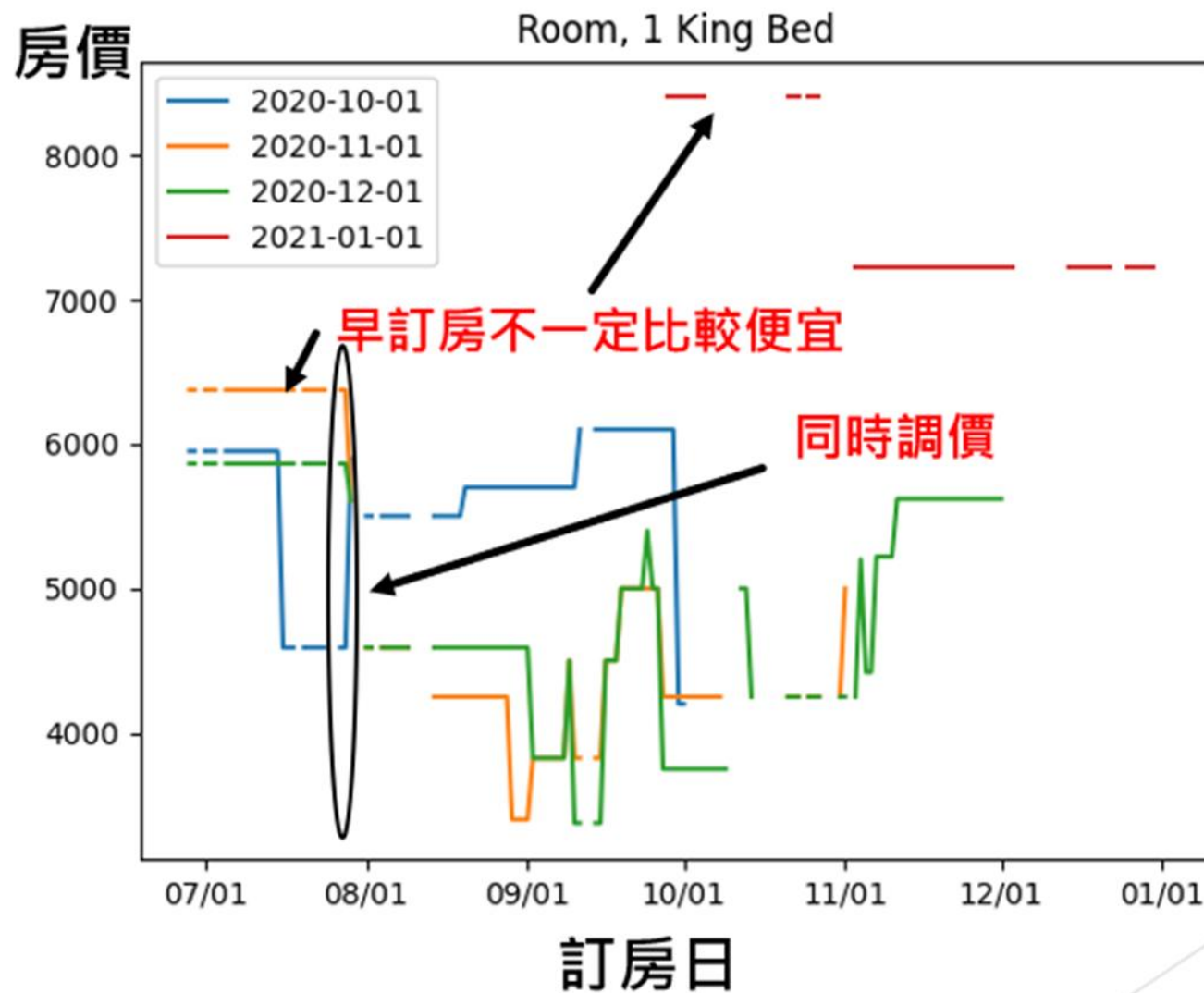
入住日對房價的影響

房價

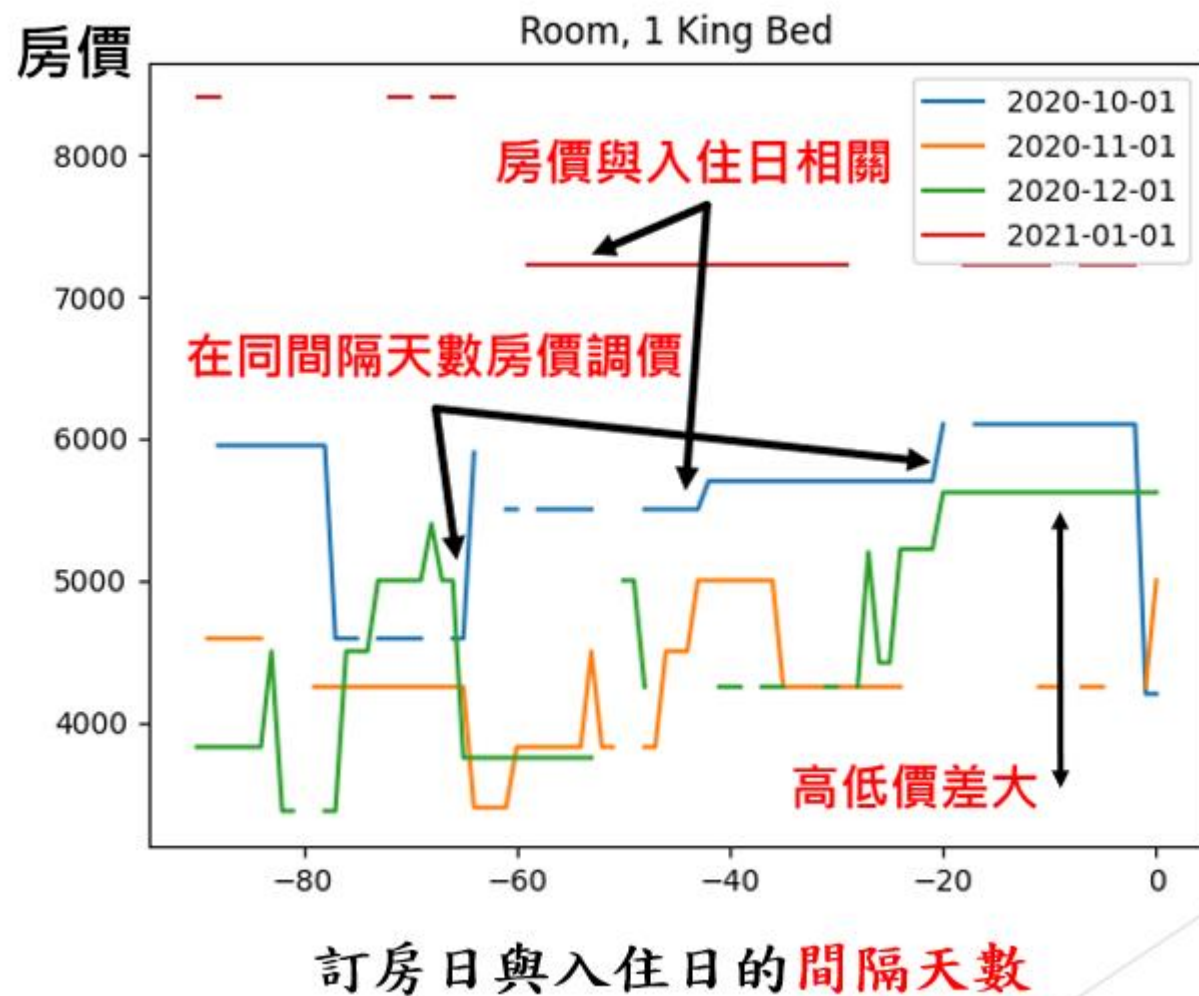
Room, 1 King Bed



訂房日對房價的影響



間隔天數對房價的影響

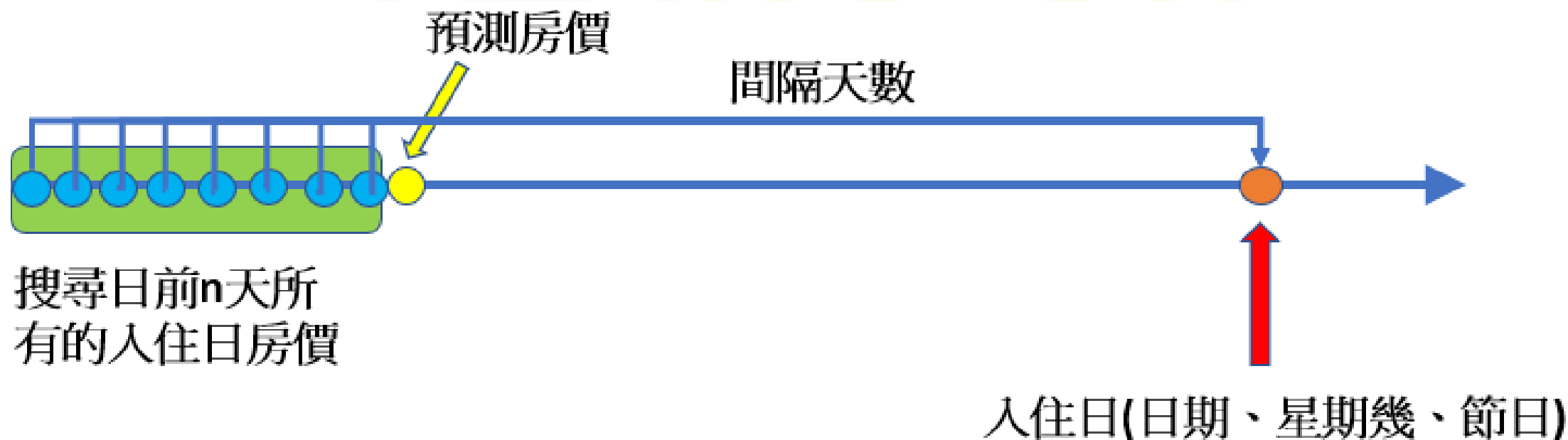


建構模型

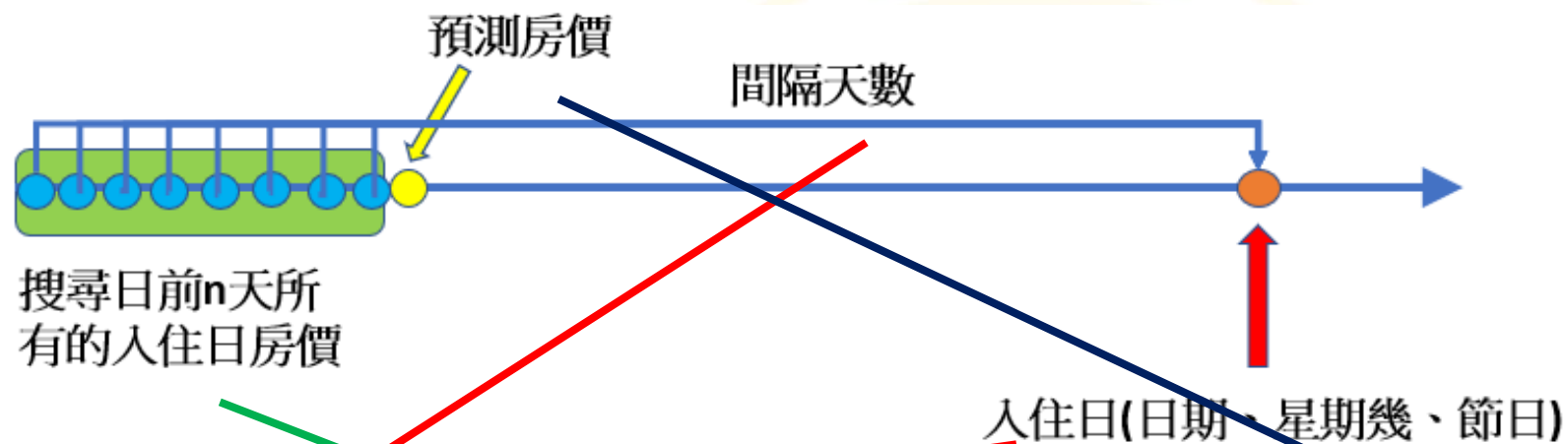
預測模型概念



預測模型目的



模型輸入、輸出資料



$$\mathbf{X} = \begin{bmatrix} \vec{x}_{1,1} & \dots & \vec{x}_{1,j} & \dots & \vec{x}_{1,n} & \vec{w}_1 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ \vec{x}_{M,1} & \dots & \vec{x}_{M,j} & \dots & \vec{x}_{M,n} & \vec{w}_M \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_M \end{bmatrix}$$

建構模型流程

資料歸一化

選定模型

選定歷史
資料天數

實際預測

資料歸一化

	1/1	1/2	1/3	1/4	1/5
Room A Price(\$)	2000	2200	2400	2600	2800
	1/1	1/2	1/3	1/4	1/5
Room B Price(\$)	3600	4000	4400	4800	5200



先找各房型中的最大最小值，進行數據的縮放

	1/1	1/2	1/3	1/4	1/5
Room A Price(\$)	0	0.25	0.5	0.75	1
	1/1	1/2	1/3	1/4	1/5
Room B Price(\$)	0	0.25	0.5	0.75	1

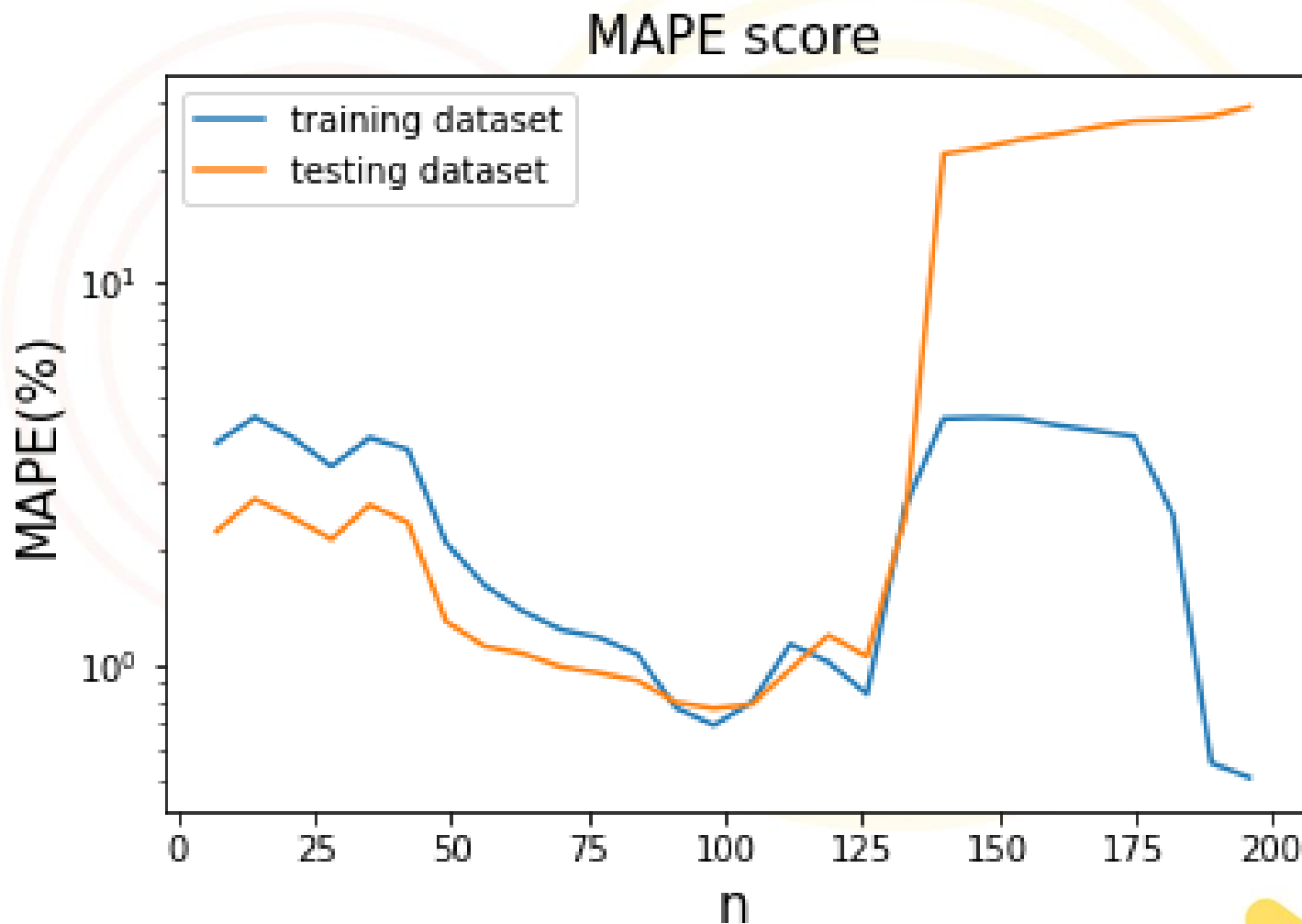
機器學習模型

- Random Forest：由決策樹 (Decision Tree) 所組成的演算法。簡單來說，此演算法可以想成是由許多人進行投票，並以多數決來得到最終結果。
- XGBoost：它是近年機器學習比賽常使用且效果不錯的演算法。而且它的運行速度也較其他方法更快、更有效率。

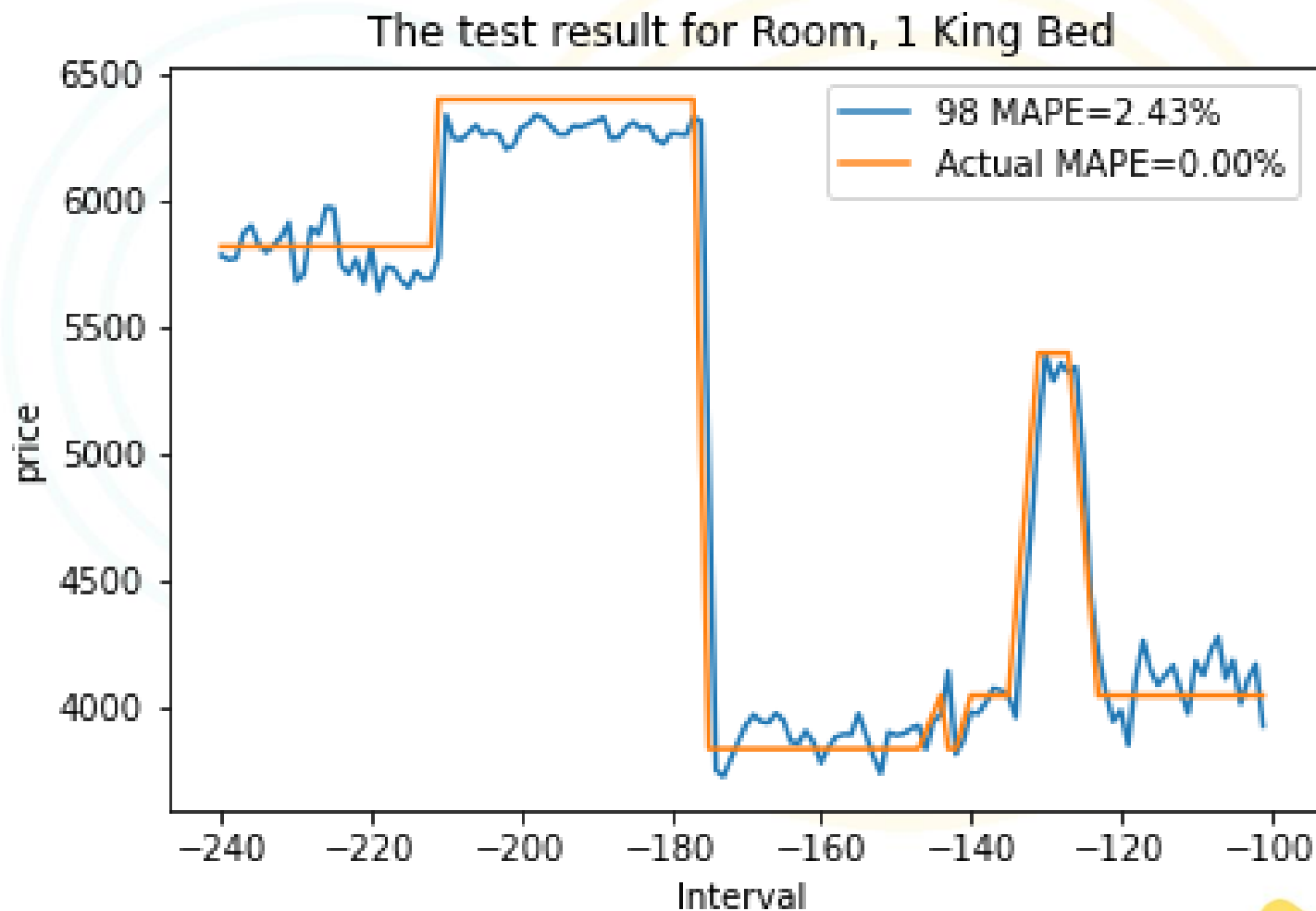
統計回歸模型

- Linear Regression：線性回歸，最基本的回歸模型
- Ridge Regression：改良版的線性回歸模型，可以減少模型的複雜性和防止過度擬合。

選定適當的歷史資料天數

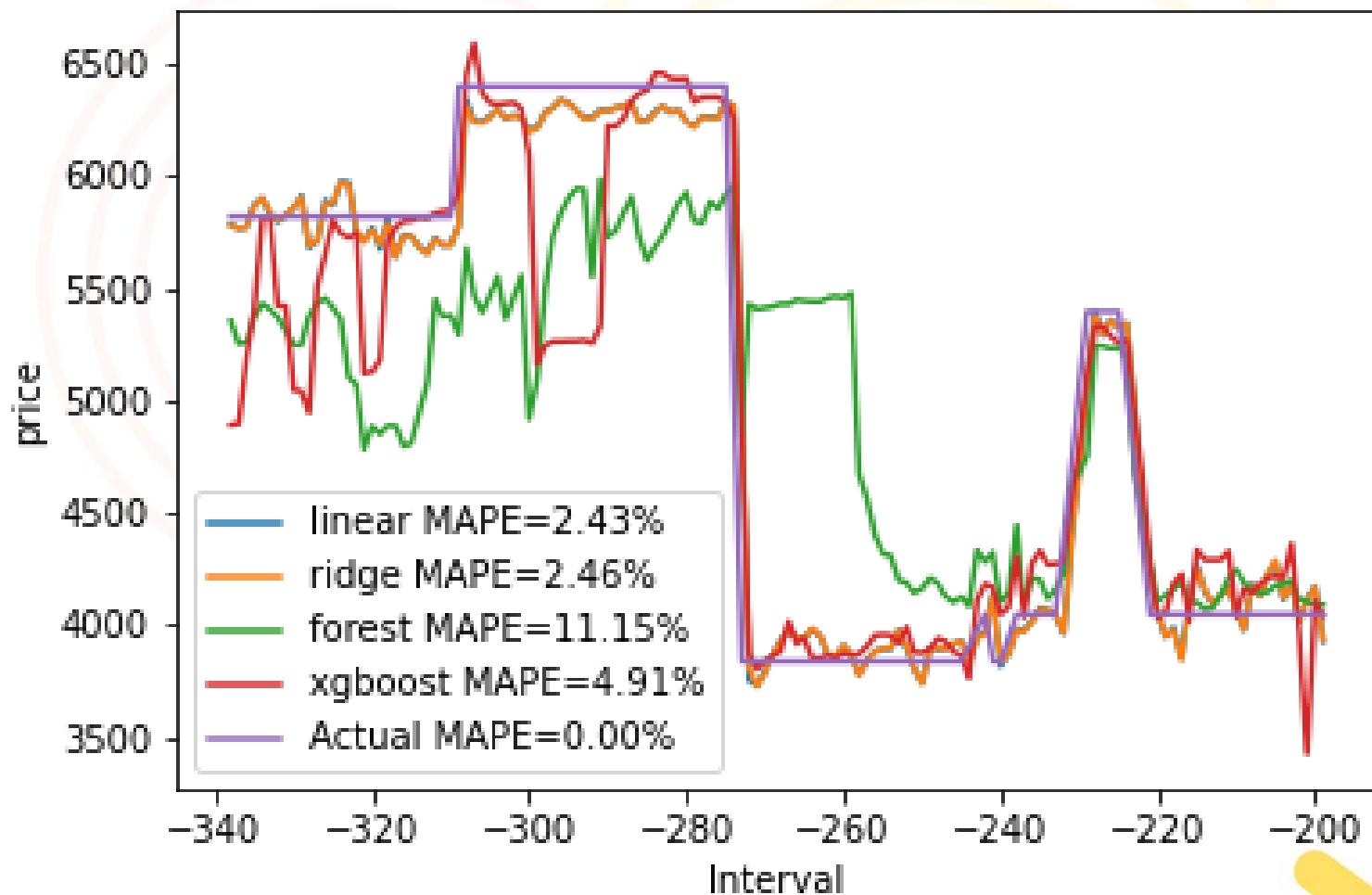


實際預測(針對6/1房價預測)



預測結果(針對6/1房價預測)

The test result for Room, 1 King Bed



結論與成果

1

透過視覺化分析，找出入住日、訂房日、間隔天數等等都是影響房價調價策略的重要變因。

2

目前針對單日預測的MAPE可達到僅0.75%

參考資料

1. T. Larrieu, "Pricing strategies in online market places and Price Parity Agreements: Evidence from the hotel industry," Working Paper, July 2019
2. Yaser S. Abu-Mostafa, Malik Magdon-Ismail, and Hsuan-Tien Lin, "Learning from data," Vol. 4. New York, NY, USA:: AMLBook, 2012.

致謝

感謝國立台灣大學資訊工程系 廖世偉教授

感謝國立台灣大學資訊工程系 謝銘鋒博士生、廖羿瑋碩士生

感謝實驗室的學長姐

感謝國立臺灣師範大學附屬高級中學 數學科 洪允東老師

感謝國立臺灣師範大學附屬高級中學 物理科 陳智勝老師

感謝國立臺灣師範大學附屬高級中學 科學班 陳昭錦主任