

## 一、EDA

### model\_and\_EDA/其他 EDA 分析/全域分析/人口流入與房價散布圖/05\_人口\_vs\_房價.png

這張圖呈現的是人口淨流入與六都房價的關係，X 軸為人口淨流入，右側為淨移入（人口增加），左側為淨移出（人口減少），Y 軸為平均房價（每坪萬元），表示該城市當月的平均房價水準，黑色實線為回歸線，顯示整體資料的線性關係趨勢。黑色回歸線呈上升趨勢，顯示人口淨流入與房價呈正相關，在城市人口持續移入的情況下，房價有較高的水準。

台北市為明顯高價區，多數點位於房價 60 萬以上區間，即便部分月份人口淨流入不顯著，仍維持高房價。中南部與桃園房價集中在中低價帶，即使有些月份人口淨流入為正，房價仍集中於 20~35 萬區間，顯示這些地區房價彈性較低。

### model\_and\_EDA/其他 EDA 分析/全域分析/六都房價排名變化/六都房價排名變化\_折線圖.png

這張圖呈現六都近十年來的房價排名變動。台北市房價穩居第一，新北第二，而桃園台中市競爭第三，南部兩市則長期房價較低，2018 年臺北市在 4 月的房價排名出現下滑，推論當月台北市房屋交易多為小坪數、或屋齡較高、地段較差的物件，拉低了當月的平均單價。臺中市當月可能成交了較多高價住宅（如七期重劃區等），推升平均房價。

### model\_and\_EDA/其他 EDA 分析/全域分析/六都整體平均房價排名/11\_房價排名 2.png

從六都整體平均房價來看，台北市以接近 70 萬/坪的價格居於領先地位，其次為新北、台中和桃園，而高雄與台南則維持相對低價水準。

### model\_and\_EDA/其他 EDA 分析/全域分析/建模分析 (ARIMA) /12\_殘差異常次數 2.png

從各都 ARIMA 殘差異常次數統計可發現，臺南與臺北市的預測表現最不穩定，異常月份次數明顯偏高，可能受到價格波動劇烈、樣本不均或外在因素（如科學園區及捷運開通等重大建設或政策）影響。相對地，臺中市殘差異常最少，預測表現最穩定，顯示其房價趨勢較為平順，ARIMA 模型擬合良好。

### model\_and\_EDA/其他 EDA 分析/全域分析/建模分析 (ARIMA) /ARIMA\_vs\_Naive\_RMSE.csv

為驗證 ARIMA 模型在房價預測上的有效性，我們進一步與 Naive 模型進行比較，並採用 Diebold-Mariano 檢定檢視兩者在預測誤差上的顯著差異。結果顯示，在臺南市、桃園市與高雄市，ARIMA 模型的預測誤差 (RMSE) 顯著低於 Naive 模型，呈現統計上顯著改善。而臺北市亦接近顯著水準，顯示其具潛在優勢。整體而言，ARIMA 模型在多數城市展現出較佳的預測表現，特別是在房價變動趨勢較穩定的區域。

### model\_and\_EDA/其他 EDA 分析/全域分析/房價回歸模型統計摘要

住宅\_平均 =  $\beta_0 + \beta_1 \text{人口淨流入} + \beta_2 \text{CPI} + \beta_3 \text{貸款利率} + \beta_4 \text{年所得} + \beta_5 \text{失業率} + \varepsilon$

### model\_and\_EDA/其他 EDA 分析/全域分析/殘差異常

殘差最大的是「台中市 2018/4」，模型預測誤差最大。其他多筆落在台北市，表示此城市在一些月份的房價變動值異常，模型難以準確預測。

### model\_and\_EDA/其他 EDA 分析/全域分析/06\_失業率\_vs\_房價 2.png

本圖檢視失業率與房價的整體關聯性，X 軸為失業率，Y 軸為平均房價。趨勢線呈現微幅下降，顯示在高失業率下，平均房價略有下降趨勢。同一失業率下，不同城市的房價落差大，顯示房價受城市本身因素主導。而將失業率離散化處理後，我們觀察到不同失業率區間的迴歸係數呈現方向不一。其中只有 3.6% 與 3.9% 顯著，其餘不顯著，代表失業率與房價的關聯性不穩定，也有可能為分類太細造成的結果波動，此外，失業率若為缺值也會顯著，會對模型統計有影響。

### model\_and\_EDA/其他 EDA 分析/全域分析/07\_變化率\_vs\_變化率.png

此圖呈現房價與利率的變化率關係，X 軸為利率變化率，Y 軸為房價變化率，資料點廣泛分布於正負兩側，無明顯線性趨勢，表示利率變動對短期房價漲跌的影響並不穩定。在多數經濟模型中，利率被視為影響購屋成本的關鍵因素，理應與房價呈現負相關。然而本圖顯示，實務上六都的房價對利率變動的即時反應極不穩定，可能因房市價格黏性高、調整延遲，或利率非房市唯一驅動因子所致。

### model\_and\_EDA/其他 EDA 分析/全域分析/08\_平均房價時序 2.png

此圖為六都房價平均時序趨勢圖，2014–2018 前期房價略呈盤整，而盤整通常代表市場處於觀望、供需平衡，2018 年出現一筆極端高峰與低谷，初步判斷為資料異常或個別高價交易影響，需進一步檢查，2019 年後房價開始穩定上升。

### model\_and\_EDA/其他 EDA 分析/全域分析/09\_變化率\_vs\_淨流入變化率 2.png

本圖呈現房價變化率與人口淨流入變化率之關係，X 軸為淨流入變化率，Y 軸為房價變化率，資料點呈現無明顯線性趨勢、也無特定集中趨勢，多數點數分布於中心四象限附近，顯示房價漲跌與短期人口淨流動變化幅度關聯性極低。

## 二、回歸模型分析

$$\text{住宅\_平均} = \beta_0 + \beta_1 \text{人口淨流入} + \beta_2 \text{CPI} + \beta_3 \text{貸款利率} + \beta_4 \text{年所得} + \beta_5 \text{失業率} + \varepsilon$$

### model\_and\_EDA/迴歸分析/VIF\_結果.csv

### model\_and\_EDA/迴歸分析/相關係數熱圖.png

在建立房價的多元線性模型後，為檢查模型中解釋變數是否存在多重共線性問題，我們結合 VIF (變異膨脹因子) 與相關係數熱圖指標進行分析，以評估模型中各解釋變數之間是否存在多重共線性問題。

所有變數的 VIF 值皆遠低於常用風險門檻 (如 5 或 10)，顯示模型中的各項解釋變數間沒有明顯的共線性問題。因此，我們可以合理信任迴歸係數的估計結果，模型在穩定性與解釋力方面表現良好。從熱圖觀察可知，CPI 與失業率之間呈中度負相關，人口淨流入與年所得之間則稍具正相關，CPI 與年所得也顯示溫和的正相關，其餘變數間則無顯著線性關係。整體而言，模型變數之間資訊重疊度低，適合作為多元迴歸解釋基礎。

### **model\_and\_EDA/迴歸分析/殘差診斷圖\_三合一.png**

(a) 殘差 vs. 預測值(檢查線性與等變異性):點並未呈現隨機分布且並未在0處附近均勻散布,高預測值處的殘差波動較大,存在異質性或模型線性不足的狀況,後續將以 log-transform 做處理。

(b) Q-Q 圖(檢查殘差常態性):中間部分大致落在紅線上,但極端處有偏離,表示殘差接近常態但存在少數離群值或偏態。

(c) 殘差 vs. 順序(檢查自相關):出現週期性波動與高頻震盪,表示殘差有時間上的自相關(ACF),可能代表某些時間序列特性尚未被模型捕捉。

### **model\_and\_EDA/迴歸分析/Durbin\_Watson.txt**

### **model\_and\_EDA/迴歸分析/Breusch\_Pagan.txt**

為檢驗模型殘差是否滿足獨立性假設,我們執行 Durbin-Watson 檢定,結果顯示自相關係數為 0.73, Durbin-Watson 統計量為 0.54,且 p 值接近 0,顯著拒絕虛無假設,表示模型殘差存在明顯的正向自相關。此現象暗示模型可能未能充分捕捉資料的時間結構。

為檢驗模型是否符合等變異性假設,我們進行 Breusch-Pagan 檢定。結果顯示檢定統計量為 11.343,自由度為 5, p 值為 0.045,顯著拒絕虛無假設,表示模型殘差存在異質性現象。

### **model\_and\_EDA/迴歸分析/迴歸模型比較.csv**

### **model\_and\_EDA/迴歸分析/Full\_vs\_Opt\_ANOVA.txt**

### **model\_and\_EDA/迴歸分析/Null\_vs\_Full\_ANOVA.txt**

### **model\_and\_EDA/迴歸分析/lm\_5fold\_CV.txt**

加入 log 變數與交互作用後建立優化模型,並與原模型比較 Adj R<sup>2</sup> 與 AIC,根據 Adjusted R<sup>2</sup> 與 AIC 的比較結果,原模型表現優於優化模型。根據 ANOVA,優化模型的 RSS 明顯上升 F 檢定的 Sum of Sq 為負,代表新模型的解釋力不但未提升,反而下降。原模型的 Adjusted R<sup>2</sup> 達 0.899,顯示具有較高的解釋力,其 AIC 為 4487.30,亦顯著低於優化模型的 4871.33。為驗證所建模型的整體解釋力,我們將其與僅含截距的 null model 進行比較。結果顯示,Full Model 顯著降低殘差平方和, F 檢定值高達 1262.7, p 值 < 2.2e-16,顯示模型中各變數對房價具有整體解釋力,支持變數納入的合理性。為進一步檢驗模型的穩健性與預測能力,本研究採用 5-fold 交叉驗證進行評估。結果顯示模型在不同資料切割下仍具一致性,平均 RMSE 為 5.72、R<sup>2</sup> 為 0.900,表明模型能穩定解釋房價約九成的變異,並具備良好預測準確度。因此保留原本的線性模型並在接下來使用進階迴歸模型。

### **model\_and\_EDA/迴歸分析/lm\_full\_RobustSE.txt**

為修正模型中因殘差異質性所造成的估計偏誤,我們採用 robust 標準誤,對迴歸模型進行強化。結果顯示主要變數如人口淨流入、年所得與貸款利率為高度顯著,表示其對房價的影響具穩定性與解釋力。CPI 為顯著而與失業率則為邊緣顯著。

### **model\_and\_EDA/迴歸分析/FE\_city\_coeff.csv**

### **model\_and\_EDA/迴歸分析/FE\_time\_coeff.csv**

在加入城市與月份固定效果後，模型能更準確地控制區域間不可觀測的差異與時間因素。縣市的固定效果係數顯示，臺北市的係數高達 59.65，遠高於其他城市，表示即使在控制 CPI、所得等變數後，其房價仍有高度結構性溢價，新北、桃園與臺中則處於中段價位，固定效果在 15–20 間，高雄與臺南則偏低，顯示南部城市即使在解釋變數相同情況下，仍有房價基期較低的結構性落差。

月份固定效果顯示，2014–2017 年房市相對平穩，固定效果波動小，顯示房價呈盤整態勢，2018 年起市場回溫，係數逐步上升，2020–2022 年固定效果大幅提升，反映房價快速成長，可能受疫情後低利與資金寬鬆影響。2023 年部分月份固定效果急遽回落。

#### **model\_and\_EDA/迴歸分析/gls\_AR1\_summary.txt**

為修正迴歸模型殘差的時間自相關問題，我們使用 GLS 模型並設定 AR(1) 結構 ( $\Phi = 0.93$ )，以符合先前 Durbin–Watson 檢定中發現的高自相關特性，針對每個城市的時間序列做殘差修正。修正後的模型顯示，人口淨流入、CPI 與年所得三項變數對房價仍具有顯著正向影響，尤其 CPI 的係數由負轉正，反映原模型中可能存在的估計偏誤已被修正。貸款利率與失業率則因誤差調整而變得不顯著，顯示其在控制時間結構後的影響力有限。模型的 AIC 降至 3787，較原模型 4487 更佳，表示整體擬合表現與資訊損失均獲得改善，殘差標準誤亦降低，標準化殘差分布良好。

#### **model\_and\_EDA/迴歸分析/WaldTest\_城市斜率差.txt**

傳統的 Panel Two-way Fixed Effects 模型假設各城市僅在固定效果(截距)上存在異質性，變數斜率則為一致。為檢驗此限制是否合理，我們進一步引入「人口淨流入 × 城市」交互項，並執行聯合線性假設檢定。結果顯示 F 值為 2.72 ( $p = 0.019$ )，顯著拒絕斜率一致的假設，表示不同城市中，人口淨流入對房價的影響力存在結構性差異。因此，在 panel 資料結構下考量斜率異質性是有必要的。

#### **model\_and\_EDA/迴歸分析/城市斜率對比.png**

為更進一步檢視不同城市對人口淨流入變數的反應是否一致，我們估計並視覺化各城市的交互斜率(人口淨流入 × 城市)及其 95% 信賴區間。新北市與臺中市的斜率顯著為正，表示人口移入對房價有明顯推升效果，臺北市與高雄市，斜率落在信賴區間內，但不顯著偏離 0，表示效果較弱或不穩定，臺南市，斜率估計值最小且信賴區間極廣，代表估計不確定性高，反映其房價對人口變動敏感度最低或數據變異大。

#### **model\_and\_EDA/迴歸分析/模型比較\_含FE\_GLS.csv**

我們彙整三種主要模型之表現指標。結果顯示，加入城市與月份固定效果的模型 ( $lm\_FE$ ) 表現最佳，具備最高的 Adjusted  $R^2$  (0.974) 與最低的 AIC (3542.45)，顯示控制不可觀測異質性後能有效提升模型精準度。相較之下，僅進行多元線性迴歸 ( $lm\_full$ ) 雖具有一定解釋力 (Adjusted  $R^2 = 0.899$ )，但 AIC 達 4487，資訊損失大。而加入 AR(1) 結構的 GLS 模型 ( $gl\_AR1$ ) 在未使用固定效果的前提下，仍能顯著降低 AIC 至 3787，顯示修正時間自相關對提升模型穩定性亦具效果。綜合而言，控制固定效果與自相關結構皆為提升模型可信度的重要手段。

#### **model\_and\_EDA/迴歸分析/CV\_metrics\_comparison.csv**

為評估各模型對房價預測的表現，我們對基準模型 ( $lm\_full$ )、優化模型 ( $lm\_opt$ ) 與城市 + 月份固定效果模型 ( $lm\_FE$ ) 進行 5-fold 交叉驗證。結果顯示， $lm\_FE$  在 RMSE (

2.72)、MAE(1.59)與  $R^2$ (0.974)三項指標皆表現最優，顯著優於其餘模型。此顯示控制城市與時間異質性有助提升模型解釋力與預測準確度。因此，lm\_FE 被選為本研究的最終模型。

### 三、區域分析

除了解釋變數對房價的影響，我們亦利用 ARIMA 模型進行未來 5 年的房價趨勢預測。ARIMA 屬於純時間序列預測方法，僅依據歷史房價進行建模，因此雖無法解釋背後機制，但能反映市場自身的趨勢動能與週期性變化。與前述迴歸模型不同，ARIMA 無法納入如 CPI、人口等外生解釋變數，因此兩種模型應互為補充，一者預測未來趨勢，一者剖析過去關係。

model\_and\_EDA/其他 EDA 分析/全域分析/建模分析(ARIMA)/10\_預測誤差異常.png

model\_and\_EDA/其他 EDA 分析/全域分析/建模分析(ARIMA)/12\_殘差異常次數2.png

為評估 ARIMA 模型在不同城市的預測能力，我們針對每個城市分別計算 RMSE 值(Root Mean Square Error)，並進行異常值標示。結果顯示，臺中市的 RMSE 明顯高於其他城市，為唯一被標記為異常的案例，可能代表該城市房價變動模式較難以單純的時間序列建模掌握。其他城市如臺北與新北雖 RMSE 較高，但仍在可接受範圍內；而桃園、高雄與臺南則表現相對穩定，顯示其房價變動具預測性。

為觀察模型預測是否在特定城市或時間點出現重大偏誤，我們統計了每座城市中 ARIMA 模型預測殘差超出兩倍標準差的次數，作為異常點的判斷標準。結果顯示，臺南市與臺北市的異常次數明顯偏高，代表這些城市在多個月份中出現過顯著預測誤差。相比之下，臺中市雖在 RMSE 指標上表現最差，但異常次數較少，顯示其預測偏誤為長期累積的小幅偏離，而非集中在特定月份的失準。

### 四、房價上升但人口負成長檢測

model\_and\_EDA/其他 EDA 分析/區域分析/臺南市/01\_房價\_vs\_人口.png

從此圖可見，臺南市在過去十年間，房價呈現穩定且明顯的上升趨勢，由約每坪 10 萬元成長至 25 萬元以上，近乎翻倍。而與此同時，人口淨流入幾乎長期處於負值區間，代表人口持續外移或流失，顯示出房價與人口趨勢的高度背離。臺南市為 ARIMA 預測異常最多的城市，雖然 RMSE 整體不高，但反覆出現嚴重預測失準且人口對房價的迴歸斜率不顯著。臺南市是六都中最具代表性的價格逆勢上升城市。可能反映以下因素：區段開發或特定建設(如沙崙智慧綠能科學城)吸引投資性購屋、本地購屋需求結構改變如年輕人口外移、長輩置產、價格受到炒作、外地資金或特定投資行為拉抬。

model\_and\_EDA/其他 EDA 分析/區域分析/臺北市/01\_房價\_vs\_人口.png

從圖中可見，臺北市的房價(中位數)在整體趨勢上持續上升，儘管中間曾在 2016–2018 年間略有波動，但自 2020 年起再度攀升至每坪約 75 萬元的歷史高點。

與此同時，人口淨流入並未同步增長，甚至有顯著下降或波動現象。雖然部分時期如 2017–2019 年間出現正向流入，但整體趨勢顯示臺北市人口移動極不穩定，並非穩定成長型都市。房價的上漲並非受到人口支撐，而更可能與資金流動、土地有限、投資性購屋、豪宅市場運作等因素有關。模型分析亦指出，臺北市在 ARIMA 模型中殘差異常次數居高，代表預測難度高，在多元迴歸中，房價對人口流入的斜率影響相對小或不顯著，進一步證實其價格形成與人口脫鉤的現象。

臺北市是一個高房價常態化且與人口流動脫鉤的案例。即便人口未明顯增加，房價仍持續攀升，顯示房價受限於供給、市場結構與政策制度，而非純粹由需求驅動。

## 五、成果總結與潛在政策建議

### 成果總結

我們針對臺灣六都房價變動進行多角度量化分析，結合房價資料與人口流動、CPI、年所得、貸款利率、失業率等多項社經變數，綜合時間序列模型 (ARIMA)、多元線性回歸與 panel 固定效果模型進行比較。

主要發現如下：房價與人口流動整體呈現正向關係，但不同城市間程度不一。臺南市與臺中市呈現人口淨流出但房價持續上升的價格逆勢上升現象。臺北市雖人口波動劇烈，房價仍長期高位盤整，顯示房價脫鉤於實質需求，與資本市場、土地稀缺等因素關聯更大。變異膨脹因子 (VIF) 與相關係數熱圖顯示模型整體無明顯共線性問題，支持各變數具獨立解釋力。Robust SE、GLS 與 Panel Two-way FE 模型進一步提升模型穩健性，也顯示「城市間差異」與「月份固定效果」確實顯著存在。

### 潛在政策建議

針對本研究觀察到的結構性價格脫鉤現象與城市間差異，提出以下政策方向：

#### 1. 關注房價與實質需求脫鉤城市

對於如臺南等人口未明顯成長卻房價劇升的城市，應加強投資性購屋審慎評估機制，例如強化實價登錄透明度、針對短期持有課徵更高稅負。

#### 2. 建立分區式房價預警系統

結合 ARIMA 與殘差異常偵測結果，可建立以 RMSE、殘差次數為基準的異常警示系統，提供政府做為預警房市過熱的重要依據。

#### 3. 導入多元解釋變數於都市發展評估

現行政策仍以人口與供需為房價評估主軸，建議納入如「資金流動」、「基礎建設預期」、「公共運輸建設密度」等非人口因素，以因應脫鉤型房市結構。

#### 4. 都市住房政策差異化

臺北市等長期高房價城市應重視「租售失衡」、「青年購屋困難」等結構性問題，推動包租代管與非自住用房源管制，而中南部則應關注區段推案過熱與產業支撐不足的風險。