

Airbnb Taipei

訂價預測模型與策略優化

Group G | 陳郁婷 楊喆東 徐尚淵 楊喻妃



Agenda



- 01 專案背景
- 02 資料前處理
- 03 重要變數挑選
- 04 模型建立
- 05 結果分析
- 06 展示影片

Agenda

01 專案背景

02 資料前處理

03 重要變數挑選

04 模型建立

05 結果分析

06 展示影片

以使用者角度拆解 Airbnb 潛在痛點



房客

- 缺乏議價空間
- 無法確認物件資訊真實性

房東

- 競品資訊難以搜集
- 訂價策略缺乏數據輔佐

由於房東在交易中較為主動、具有訂價能力和較多的選擇空間，因此我們決定聚焦其痛點上，藉由提升訂價準確度來提高成交機會，並增加 Airbnb 來自房客與房東的平台使用費營收。

專案目標

協助 Airbnb 房東優化訂價策略

執行方法

利用機器學習模型以物件資訊作為特徵預測房價

資料來源

Inside Airbnb - Taipei, 25 September, 2023

Agenda

01 專案背景

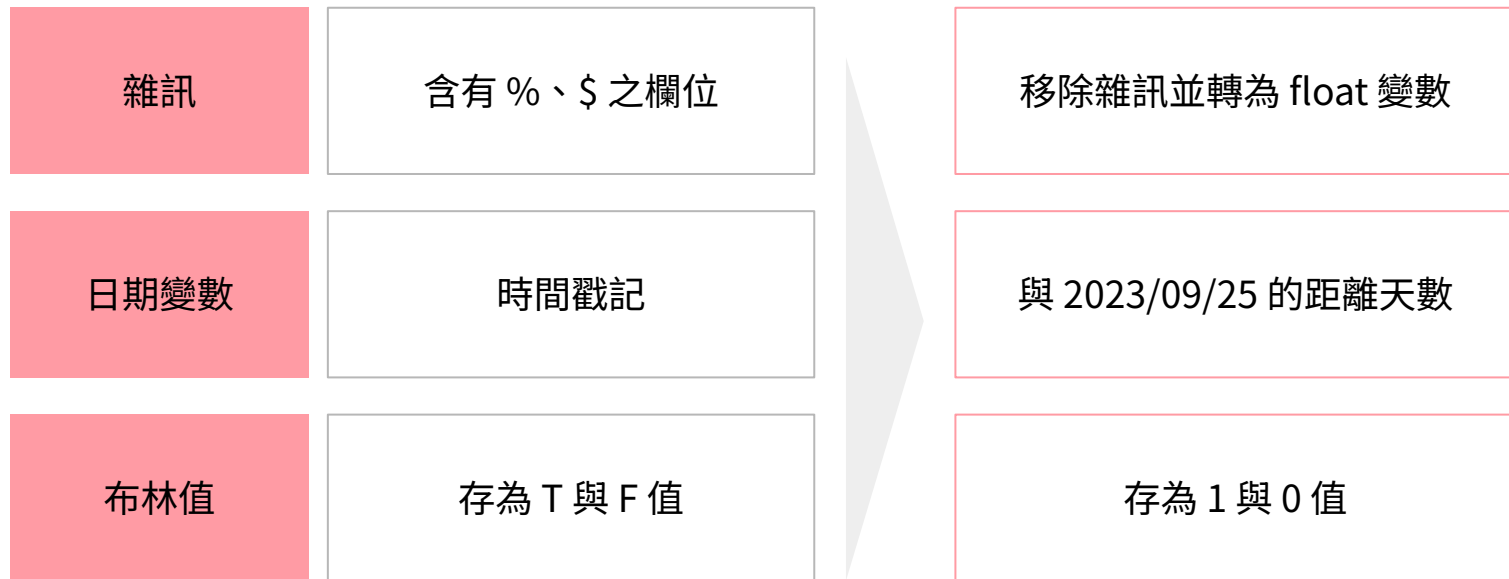
02 資料前處理

03 重要變數挑選

04 模型建立

05 結果分析

06 展示影片



不必要欄位

刪除與房價較無相關的欄位
如 scrape_id、host_name 等

價格離群值

刪除房價於第 1 四分位數與第 3 四分位數
 ± 1.5 四分位距以外之物件

高相關變數

針對各組高相關變數，保留其中一項變數

類別變數

利用**眾數**填補類別變數缺失值，
包含 room_type、property_type 等

數值變數

利用 **MICE** 填補數值變數缺失值，
包含 host_response_rate、price 等

bathroom_text

- 原始資料：6.5 shared baths、1 private bath 等
- 由於類別眾多且帶有複合資訊 (數量+種類)，因此將 shared bath、private bath 等作為獨立的欄位，分別儲存相對應之數量

amenities

- 將原始文字斷詞，並清洗標點符號、空格等雜訊
- 由於設備種類眾多，因此先利用人工將其分成較大的類別，如廚房用品、嬰兒用品等，再進行 One-hot encoding

其他欄位

- 針對 room_type、property_type 等種類較為單純之類別變數
- 採用 One-hot encoding

Agenda

01 專案背景

02 資料前處理

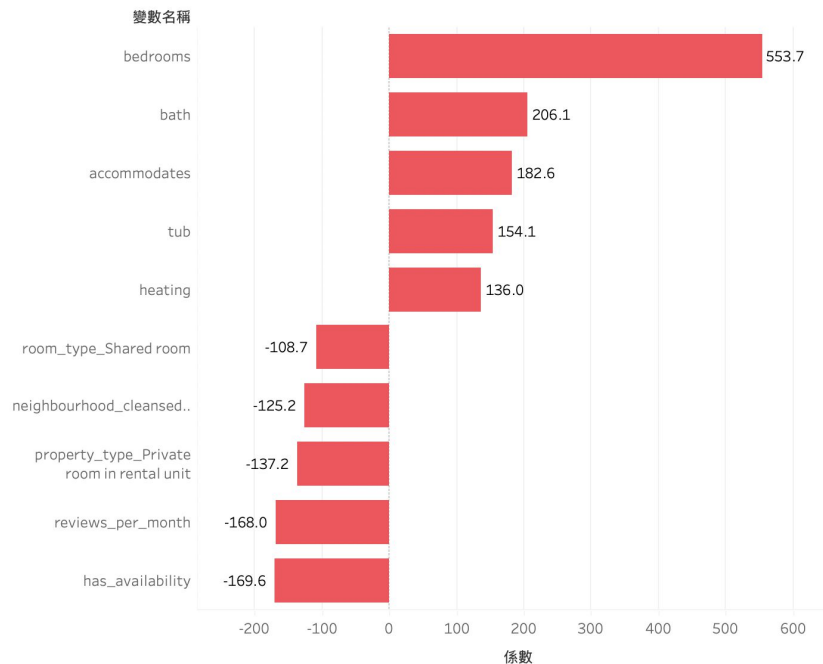
03 重要變數挑選

04 模型建立

05 結果分析

06 展示影片

利用 Elastic Net 進行特徵挑選

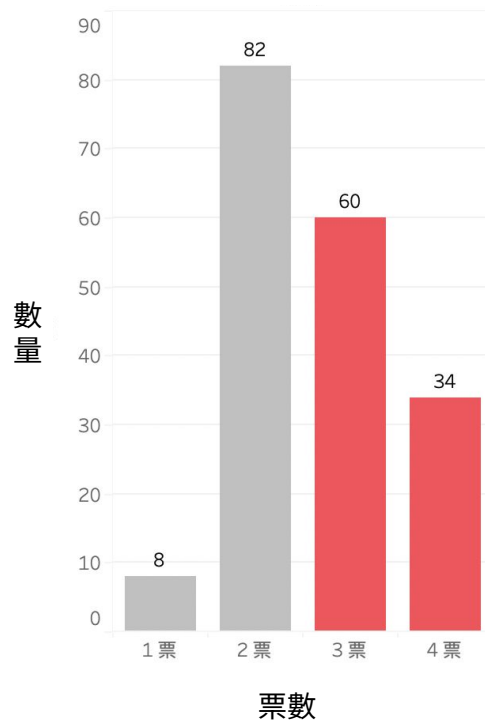


▲ Elastic Net 重要係數排序

原特徵數量 **184 個**

投票法

- Random Forest
- XGBoost
- Lasso
- Stepwise



Agenda

- 01 專案背景
- 02 資料前處理
- 03 重要變數挑選
- 04 模型建立**
- 05 結果分析
- 06 展示影片

資料集切分

64%

Training

16%

Validation

20%

Testing

模型訓練

- 針對三種重要變數資料集，實測 Linear Regression、XGBoost、SVR、Random Forest 等模型
- 針對各模型分別以 GridSearch 優化超參數，並以 K-Fold Cross Validation 避免過度擬合

結果比較

- 以 RMSE 與 Adjusted R squared 比較各模型之預測結果，選出效果最佳者

針對以 Elastic Net 所挑選出之重要變數建立模型

	RMSE	Adjusted R-Squared
Random Forest	1017.306	0.650
Linear	1255.223	0.384
SVR	1267.424	0.424
XGBoost	1275.348	0.366

針對以投票法得 4 票所挑選出之重要變數建立模型

	RMSE	Adjusted R-Squared
Random Forest	988.679	0.681
Linear	1263.997	0.427
SVR	1292.415	0.351
XGBoost	1276.693	0.444

針對以投票法得 3 和 4 票所挑選出之重要變數建立模型

	RMSE	Adjusted R-Squared
Random Forest	997.085	0.671
Linear	1350.917	0.389
SVR	1281.685	0.411
XGBoost	1276.291	0.413

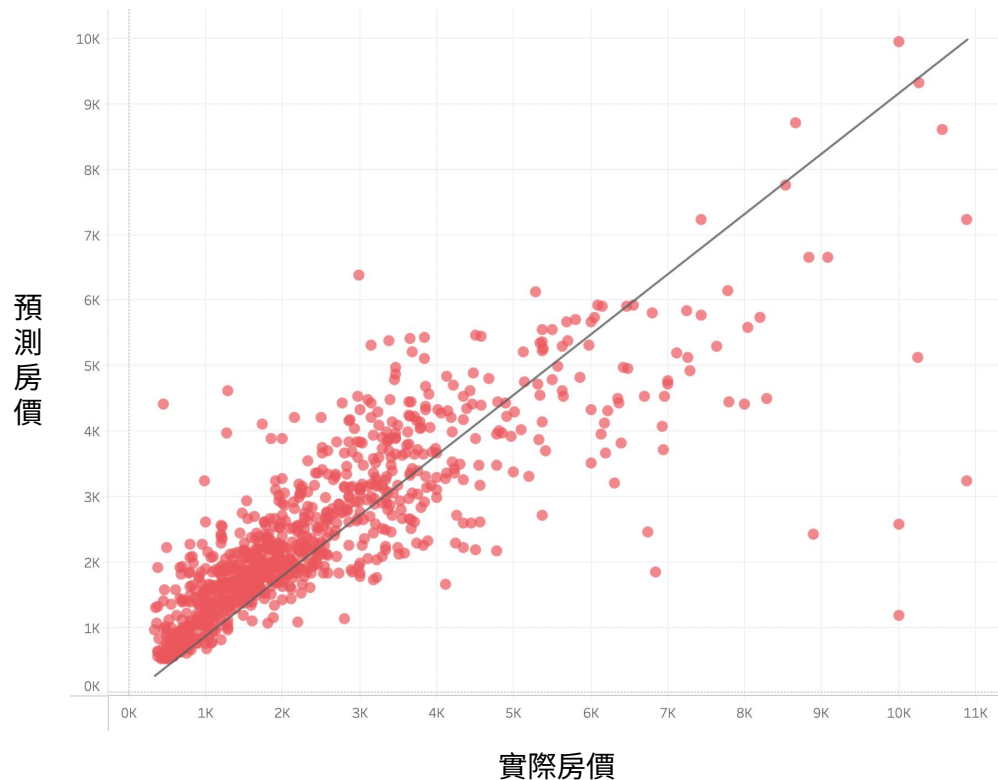
Random Forest 在各個重要特徵種類表現比較

	RMSE	Adjusted R-Squared
投票法 4 票	988.679	0.681
投票法 3 與 4 票	997.085	0.671
Elastic Net	1017.306	0.650

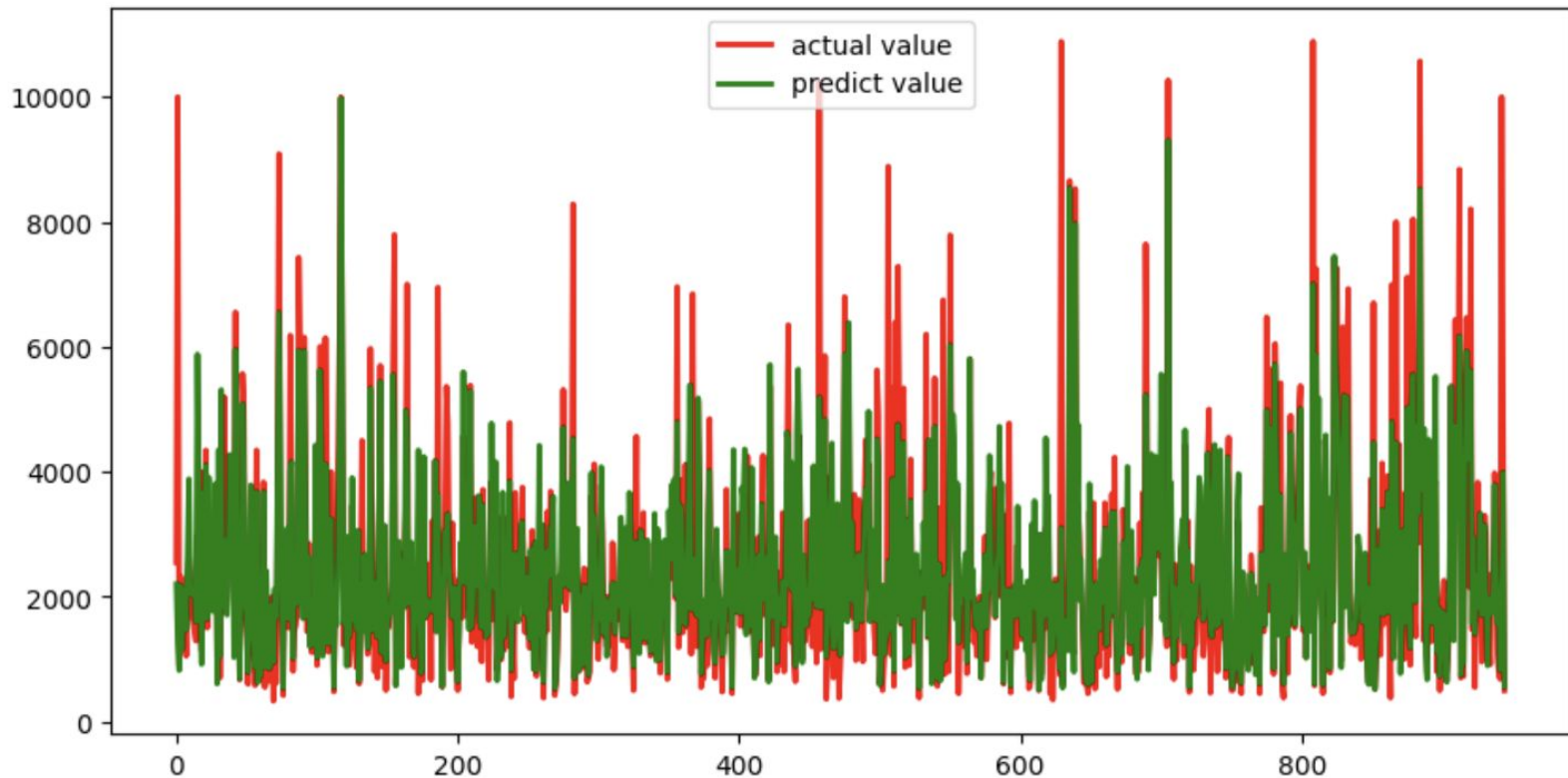
結論

Random Forest 為表現最佳之模型，其中又以使用投票法 4 票的特徵數時表現最佳。因此後續將以**投票法 4 票**之變數作為特徵，並以 **Random Forest** 作為模型進行分析。

Random Forest Regressor 模型表現



Random Forest Regressor 模型表現

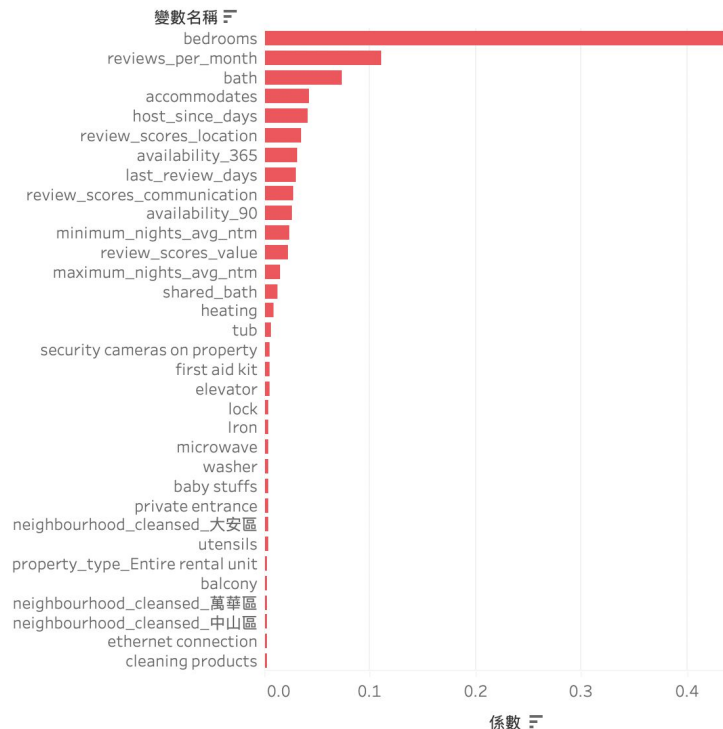


Agenda

- 01 專案背景
- 02 資料前處理
- 03 重要變數挑選
- 04 模型建立
- 05 結果分析**
- 06 展示影片

可將重要變數歸納成六大類

基礎設備	包含清潔、衛浴、嬰兒用品等
地點	大安區、中山區、萬華區
出租狀況	包含租借天數上下限等
評論	包含評論數量、分數、最新評論日期等
屋主狀況	屋主年資
房子型態	是否為包棟建築



訂價優化系統

說明

房東在刊登物件時會選擇要提供哪些資訊，而系統會為房東比較類似競品在市場中的訂價，並提供房東建議售價，以及若房東再多增加哪些資訊，即可提高房客多少的願付價格。

可解決的痛點

- 競品資訊難以搜集
- 定價策略缺乏數據輔佐

Agenda



- 01 專案背景
- 02 資料前處理
- 03 重要變數挑選
- 04 模型建立
- 05 結果分析
- 06 展示影片**



The image shows a screenshot of the Airbnb website's rental form. The form is titled "出租Airbnb 並可賺取" (Rent Airbnb and Earn) and includes a sub-header "讓我們介紹你的房源" (Let us introduce your listing). The form is divided into several sections: "一次最多能住多少人?" (How many people can stay at once?), "地點" (Location), "租金" (Rent), and "新選擇" (New Choice). The "新選擇" section includes three buttons: "租內租設備" (Rent in-house equipment), "獨立專用車位" (Independent dedicated parking space), and "其他選擇" (Other choices). The form also features a "Airbnb 入門" (Airbnb Intro) button in the top right corner and a "下一步" (Next Step) button at the bottom right.

讓我們介紹你的房源

一次最多能住多少人?

人數

地點

租金

新選擇

租內租設備

獨立專用車位

其他選擇

Airbnb 入門

下一步

Thanks for listening.