**北大 x 好域**

時間：2020/2/13（四） 14:00

參與人：沈易星學長、林奕銜、蕭琮峻、許程閔

上次[會議紀錄](https://docs.google.com/document/d/1PxZ2NSdUxUu5j6VUb1BGEWU1at9BNIxjFMxRiN79QlQ/edit)，[Working Process](https://docs.google.com/document/d/1AUf4gI5kxPYSEiniz-MhMeHtMWicBmLqcMjot6EHihM/edit?usp=sharing)

**目前進度：**

1. Lead\_time和「預測模型」進行交叉驗證，看兩者是否有關聯。[連結](https://drive.google.com/open?id=1PO5idTNtdLMo17SHZuB2WSjRqN0D3sYQ)  
   相關係數為 : -0.008

[lead\_time程式連結](https://drive.google.com/open?id=1sC2rraCRhLQOAPgTTynnB9beRutSzYk8)

[lead\_time資料表](https://drive.google.com/open?id=1bYBjGyOczGyViVLoHmBXvhW_VZF3y-xg)

1. space準確率在75%以下的模型，可能是under fitting的問題。
   1. 可先將模型依照準確率高低排序，並按模型數量平均分為四群，試圖找出原因。
   2. 就學理上的瞭解，將under fitting的解決方式套用在模型上，並觀察模型績效是否有提升。
2. 程式化隨目前進度持續修改及改善。[連結](https://colab.research.google.com/drive/19vzpn55u9JkY9UOkAcP9E13Uae3pSy4A)

**上次進度：**

1. 改以space作為單位進行預測，並查看準確率。（奕銜）[連結](https://drive.google.com/open?id=11ouMTHGzBtOj4B4LAnl93hXuTMMMHaBK)
2. 修正目前的Lead\_time，計算所有order的平均數與變異數，計算出每個space分別的（星期與時段）各自的lead\_time與z-score 。（程閔）
3. 修正程式，依照全部order算出平均值，將各個space依照每周7個weekdays以及每天4個時段計算出各自的mean\_lead\_time，並比照母體平均數，得出z-score。[code連結](https://drive.google.com/open?id=1L9qDaraxERCXWlamBkzXnDnCMT4RbZzW)
4. Lead\_time 和「預測模型」進行交叉驗證，看兩者是否有關聯。（資料還需要再處理）
   1. 預測每一個空間每一天的每一個時段的出租機率
5. 根據space屬性關係預測，做出目前房型的預測定價，提供決策參考。[連結](https://drive.google.com/open?id=1fHtE0-5S-R2tZdE-90vGA9VEzhHTTxzQ)

**下次嘗試：**

1. space準確率在80%以下的模型，可能是under fitting的問題。
   1. 可先將模型依照準確率高低排序，並按模型數量平均分為四群，試圖找出原因。
   2. 就學理上的瞭解，將under fitting的解決方式套用在模型上，並觀察模型績效是否有提升。
2. Leadtime跟準確率分開看。

＠前題：要先確認space的Lead time是不是在一個禮拜之外

1. Lead time前價格調漲，Lead time一個禮拜內續繼續看預測機率看看是要減價漲價，機率高：維持；機率低：降價。
2. 一個禮拜內的Lead time跟準確率要怎麼做結合。

**上次的問題討論：**

1. 價格訂定之機制
   1. 機率預測模型以達到多少 % 為出租出去的標準？（50%／75% etc.）
      1. 待討論。
   2. 預測機率、Lead\_time與**價格**的關係？
      1. 待討論。
   3. 動態定價顆粒度（時段／天／週）？
      1. 越細越好，應該要每個時段都能有自己的價格。

**未來要注意的事情：**

1. 羅吉斯回歸預測模型，ROC Curve要切的機率點可以再討論和確定。
2. 思考最適價格的預測（與當前定價比較）。
3. 可思考依據那些因素（季節、使用率）觀測市場的需求變化，可再依此做出價格調整。
4. 純粹根據房間資訊來進行預測，找出同一個locate當中，不同space的價格符不符合目前的價格模型，再進一步來進行調整。
5. 自動化檢定比較新舊資料差異（可從使用率進行比較）。

會使用到的原始資料：

* spaces.csv
* spaces\_times.csv
* spaces\_devices.csv
* orders.csv
* god\_account\_view.csv

核心程式：[連結](https://colab.research.google.com/drive/19vzpn55u9JkY9UOkAcP9E13Uae3pSy4A)

* 下列為程式會產生出的檔案

1. Phase 1: Create necessary data
   1. \* [Space\_info](https://drive.google.com/open?id=1xlW-pEEos8QoUz_YUNNQdCXpMET4XZi1) (For Generate Fake Order)
   2. \* [Real Order](https://drive.google.com/open?id=1l8ByaEIDClP0PBZ_Zb_o3FnnjbxTZd13) (For Generate Fake Order)
2. Phase 2: Generate Fake Order
   1. [Fake\_Order\_in\_Empty\_Time](https://drive.google.com/open?id=1fAhTkt9oKpUKtHgtbVj8IRz2jGrtUczz) (For c. combine)
   2. [Fake\_Order\_in\_Real\_Order\_Time](https://drive.google.com/open?id=1E8s82tr2z0uR-HP-_E0CslEA42e-KGj3) (For c. combine)
   3. \* [Fake & Real Order Combine](https://drive.google.com/open?id=165OJ3XqADkfGKW4kAcX1EtQWKOilKVEW) (For predict model, calculate lead\_time)
3. Phase 3: Setting Price
   1. [Locate\_Lead\_Time](https://drive.google.com/open?id=18ansq4XL-k3hySMV-51-njxY7mPnDUqv) (For setting Price)
   2. **Final Recommanded Price Table**