**北大 x 好域**

時間：2020/01/30（四） 14:00

參與人：陳宗天教授、沈易星學長、林奕銜、黃俊友、蕭琮峻、許程閔

上次[會議紀錄](https://docs.google.com/document/d/1uf5cI1uVQPh1xjNmfh9p4c1dzHCnLY0f3xYdUgXWTDc/edit)，[Working Process](https://docs.google.com/document/d/1AUf4gI5kxPYSEiniz-MhMeHtMWicBmLqcMjot6EHihM/edit?usp=sharing)

**問題回答：**

1. 請問excel裡面的standard\_n是什麼意思？
   1. standard\_n代表n個標準差內。
2. 感覺mean的算法感覺怪怪的？
   1. 如果看space跟locate數量的話，會忽略了訂單數量的多寡，忽略訂單數量可能會造成平均的偏誤？
      1. 訂單內的lead\_time，是對應的space總total\_lead\_time / order 數目，也就是說是平均後的資料。
   2. 請問把mean分成space＆locate的用意思？是不是可以考慮使用total\_lead\_time/ number\_of\_data算出所有訂單的mean？然後再依照這個去看是否符合常態分配？
      1. 價格主要是針對space做調整，所以主要是看space那邊的資訊。會額外多出locate是要看有沒有區域整體lead\_time偏差的問題，方便日後做決策，若是不需要的話可以考慮日後拿掉。
      2. 目前space中的lead\_time是使用這樣的算法。
3. 如果要依照space或locate看的話，感覺應該也會是this\_space\_lead\_time/ number\_of\_this\_space\_order 或 this\_locate\_lead\_time / number\_of\_this\_locate\_order，理論上應該是每個space/locate有自己的lead\_time才對？
   1. 目前是這樣，每個space或locate的lead\_time，都是平均後自己的lead\_time。
4. 請問在資料裡面，mrt之後的status＆0~19是代表什麼？
   1. status = 1 代表未出租出去的空訂單時段；status = 2 代表成功出租。
   2. 0~19 則是代表該訂單房間所擁有的設備 device\_id。
5. 這邊放的連結好像不是程式。
   1. 已修正，原先的連結應為當前程式會產生出的檔案。
   2. 程式連結應為下方單一程式化的連結。

**目前進度：**

1. 單一程式化（[連結](https://colab.research.google.com/drive/19vzpn55u9JkY9UOkAcP9E13Uae3pSy4A)）
   1. 資料處理：
      1. 資料前處理（琮峻）
         1. 刪除不必要的欄位並新增time group, weekday, is\_holiday, locate\_id（space要選用有MRT的）
         2. 合併space & order
         3. space\_device轉置
         4. 將轉置結果合併至space&order
      2. 假訂單產生（俊友）
         1. 依據Space資訊產生全時段假訂單。[Fake\_Order\_in\_Empty\_Time](https://drive.google.com/open?id=1fAhTkt9oKpUKtHgtbVj8IRz2jGrtUczz)
         2. 依據Order資訊產生實際訂單當天的空時段假訂單。[Fake\_Order\_in\_Real\_Order\_Time](https://drive.google.com/open?id=1E8s82tr2z0uR-HP-_E0CslEA42e-KGj3)
      3. 真假訂單篩選與合併。（琮峻）
   2. 模型訓練（奕銜）
      1. 將全部資料套入 Regression model 查看準確率(總體模型)。
      2. 將使用者想預測的locate的資料套入 Regression model 查看準確率(個別模型)。
      3. 看總體模型與個別模型的準確率哪個高，便使用那個當作該locate的預測模型。
   3. 價格訂定（程閔）
      1. 初步策略
2. Lead\_time 與價格訂定之間的關係與機制。
   1. 目前分析space資料發現（1, 2, 2+）個標準差內分別為 (63%, 92%, 96%) 大致上符合常態分布，因此初步制定價格調漲策略為：1個標準差內價格不更動，2個標準差內價格上下調整1成，3個標準差內價格上下調整2成，4個標準差以上價格從2成到3成之間調整（4個標準差= 2.1 ，5個標準差= 2.2 … 最高到3成）

**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。

**下次嘗試：**

1. 改以space作為單位進行預測，並查看準確率。（奕銜）
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 。（程閔）
3. Lead\_time 和「預測模型」進行交叉驗證，看兩者是否有關聯。
   1. 預測每一個空間每一天的每一個時段的出租機率
4. 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。

**上次進度：**

1. 計算房間預定的前置期：
   1. 運用訂單的Created\_at 的時間，和實際預定的日期，去計算每個space、locate的平均lead time（以小時為單位）和變異數，並且依照標準差來做定價。而若小於24小時仍尚未出租，則應額外降價來吸引顧客。

使用更新後的[資料表\_code](https://drive.google.com/open?id=18ansq4XL-k3hySMV-51-njxY7mPnDUqv)

[Space and Locate lead time](https://drive.google.com/open?id=1paEZeCNvbqPAFANCNs-_UrHnRzzxgM4O)

1. 修改過後的[模型準確率](https://drive.google.com/open?id=1GnDjmOiJb1_AdsKF2KrrAH8OOFGJF880)
   1. 平均Train Set Accuracy：82.16%，與舊資料相比增加了1.17%。
   2. 平均Test Set Accuracy：82.81%，與舊資料相比增加了2.76%。
2. Space各個屬性與＂價格訂定＂兩間之間的關係與機制：
   1. 利用線性回歸模型來預測，訓練集準確率88.05%，測試集準確率73.27%。
   2. 代表目前各個Space的價格訂定有一定標準。

**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
* orders.csv
* spaces\_devices.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**