**北大 x 好域**

時間：2019/11/28（四） 16:00

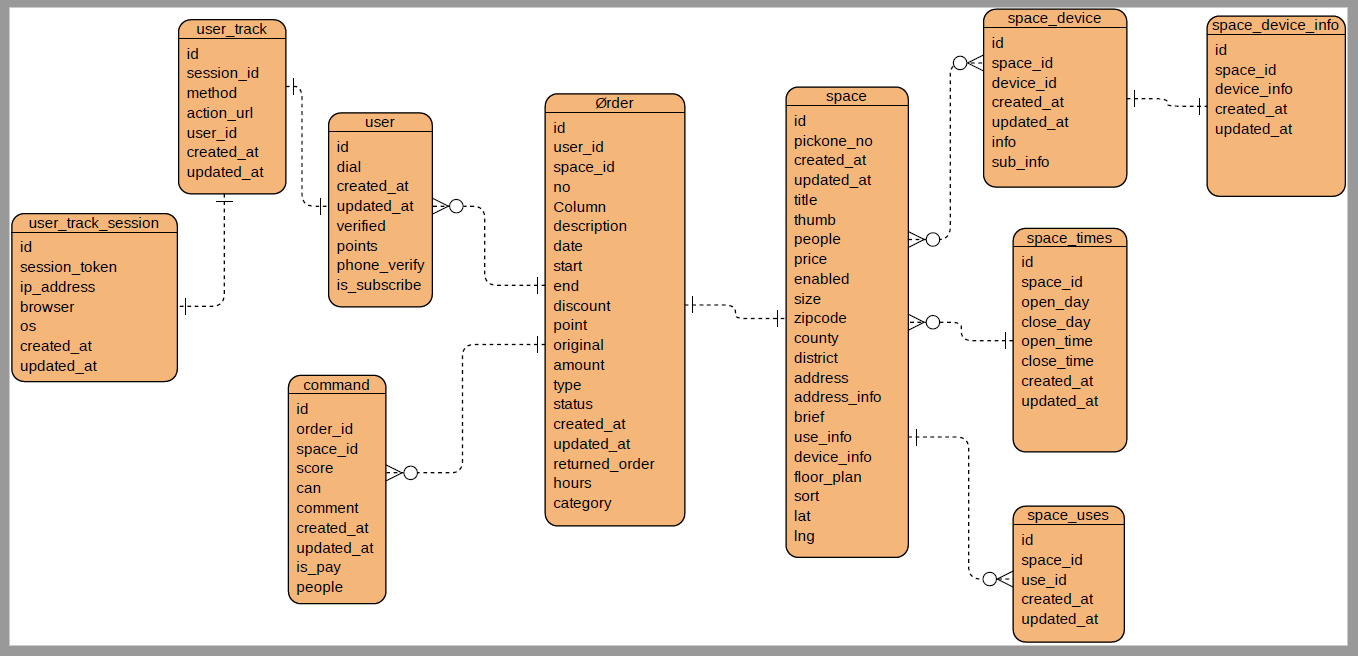
參與人：陳宗天教授、沈易星學長、林奕銜、黃俊友、蕭琮峻、許程閔、溫智皓

上次[會議紀錄](https://docs.google.com/document/d/17SsFDoFfypG5pyXJL-ZonR85R60pFe4fur3MXGaAPII/edit?usp=sharing)整理：

* 開發環境：使用 Github 建構管理，並在 Google Colab 使用 Python 開發。
* 預測模型：初步階段，使用羅吉斯回歸模型預測出租機率，檢測績效後，未來再使用其他模型進行預測並比較績效。
* 套件使用：tensorflow、sk-learn、keras或其他套件，並研究要使用哪種套件。
* 專案變數：
* 應變數：出租與否、推薦價格。
* 自變數：地點、房間、日期、是否為假日、時段、人數、坪數、型態、實際價格、季節等其他自變數，未來需進一步釐清。
* 專案目標：
* 計算空間出租機率。
* 推薦合理出租價格。
* 動態調整價格，提高空間利用率。
* 資料處理：
* 整理由好域提供的 csv 檔案，將建模所需要使用的資料，合併在同一個檔案。
* 可針對各式資料，自行做修改或應用，產生可用的新變數。

目前進度討論：

1. 初步資料前處理已完成，目前分為兩部分：
   1. Total Database using Variable: All related data.
   2. Predict Model using Variable:
      1. orders
         1. space\_id
         2. category
      2. spaces
         1. space\_id
         2. people
         3. size
         4. order\_price
         5. price
         6. zipcode
      3. New created variable
         1. is\_holiday ( 0: Mon. ~ Fri., 1: Sat. ~ Sun.+ Country Holiday )
         2. per\_person\_price ( price / people )
         3. **per\_size\_price ( price / size )**
         4. time\_group ( 1: start ~ 12:00, 2: 12:00 ~ 14:00, 3: 14:00 ~ 18:00, 4: 18:00 ~ end )
         5. weekday ( 1: Mon. .... 7: Sun. )
         6. hours ( renting hours )
2. Predict Model:
   1. 使用Sk-learn的train\_test\_split將資料集分為訓練集、開發集與測試集，比重分別是60%/20%/20%。
   2. 目前使用 Sk-learn 套件來進行預測，先建立一個可運作之模型後，未來可嘗試使用其他套件，再比較績效。
   3. 目前已可丟資料進去進行預測，但因資料處理尚未完成，跑出來的預測結果還不是我們要的目標。
3. ER model:



下一步該進行的事：

1. 製造 "無訂單時段" 的資料放進 Predict Model 的 database 。
2. 缺失值處理
3. Time\_group設置兩欄，處理跨時段的訂單
4. 待資料處理完成後，將資料丟入模型，預測空間出租機率。
5. 完善資料及模型。

目標新增：

1. 未來有新的租屋加入時，根據其房屋資訊預測其價格。
2. Compute the average time between the order finished and the actual rending time, for designing the promotion 計算付款完成時間到實際租房日的平均時間，以利促銷活動的設計。

Questions:

1. mrt columns -> future
2. Some space info is empty or weird.
   1. people column = 0 -> ignore
   2. weird space info -> delete

Work Remind:

* Even not use from the beginning, try to remain the information about orders and spaces. For later we may need it. 儘量保留與訂單、房間相關的資訊
* Establish ER model 建立ER模型
* Note down every workflows 紀錄每一個工作流程
* Train / Vaildation / Test : 60/20/20 訓練集 / 驗證集 / 測試集
* Compute the average time between the order finished and the actual rending time, for designing the promotion 計算付款完成時間到實際租房日的平均時間，以利促銷活動的設計。
* Data process needs programmability in final overall. 資料處理最後必須程式化、自動化