

資料探索作業說明

Ubike資料集 (台北市公開資料)

■ 656711筆資料、21欄位

屬性名稱	說明	屬性名稱	說明
date	日期	std.sbi	車輛數標準差
hour	時間	avg.bemp	平均空位數
sno	場站代號	max.bemp	最大空位數
sarea	場站區域	min.bemp	最小空位數
sna	場站名稱	std.bemp	空位數標準差
lat	緯度	temp	平均氣溫
lng	經度	humidity	溼度
tot	總停車格	pressure	氣壓
avg.sbi	平均車輛數	max.anemo	最大風速
max.sbi	最大車輛數	rainfall	降雨量
min.sbi	最小車輛數		

題目說明

- 請探索此資料是否存在特定樣型(pattern)或規則(rule)，並試著說明或解讀產生此樣型的原因以及可能衍生的研究議題。
- 建議做法
 - 定義每個欄位應有的衡量尺度
 - 思考欄位之間是否可組合產生有意義的新欄位
 - 畫圖、作敘述統計，檢視是否有遺漏值、離群值
 - 單變量與多變量交互檢驗，討論是否有可疑訊號，思考背後的合理性
 - 提出研究命題或假設，以及對應檢驗的方法的proposal

繳交格式

- 每組以A4打字撰寫報告，字體大小12，單行間距，頁數不限。報告第一頁統一呈現方式如以下摘要表格

編號	樣型/規則描述	提出假設/研究命題	預計檢驗方式
1			
2			
3			
...			

後續可自由排版詳述資料探索過程(包含所使用的方法與工具)，以及如何發掘各樣型/規則的過程與假設建立，若有相關參考資料請比照一般研究文獻整理列在最後。