

## D4：EDA: 欄位的資料類型介紹及處理

[PDF 下載](#)[全螢幕](#)

### Sample Code & 作業內容

A、在初步 EDA 的過程，我們無可避免會想問的問題  
不同資料類型各有多少個欄位？  
類別型欄位 (pandas 中的 object) 的類別數量？  
模型怎麼處理類別型的資料？有什麼表示方法？

B、這三個問題在參考程式碼範例(請點選下方檢視範例Day\_004\_column\_data\_type.ipynb)都會實現，第三個問題會更為複雜一些，簡單來說我們有兩種方法來處理類別型資料

- Label encoding: 把每個類別 mapping 到某個整數，不會增加新欄位
- One Hot encoding: 為每個類別新增一個欄位，用 0/1 表示是否

Label Encoding			One Hot Encoding			
Food Name	Categorical #	Calories				
Apple	1	95	Apple	Chicken	Broccoli	Calories
Chicken	2	231	1	0	0	95
Broccoli	3	50	0	1	0	231
			0	0	1	50

C、作業1：請詳細閱讀[label-encoder-vs-one-hot-encoder](#)（英文）  
D、作業2：請點檢視範例，作業請提交Day\_004\_HW.ipynb檔

[檢視範例](#)

### 參考資料

課後補充:

[Pandas 數據類型](#)

### 提交作業

請將你的作業上傳至 Github，並貼上該網網址，完成作業提交

[確定提交](#)[如何提交](#)

到 Cupoy 問答社區提問，讓教練群回答你的疑難雜症

[向專家提問](#)[如何提問](#)