

[首頁](#)[個人主頁](#)[活動資訊](#)[排名](#)[公告](#)[常見問題](#)[我要提問](#)[f](#)[張](#)[登出](#)

100 道題目 > D12 : EDA: 把連續型變數離散化

D12：EDA: 把連續型變數離散化

[PDF 下載](#)[全螢幕](#)

Sample Code & 作業內容

今日作業：

- 新增一個欄位 `customized_age_grp`，把 `age` 分為 `(0, 10]`, `(10, 20]`, `(20, 30]`, `(30, 50]`, `(50, 100]` 這五組，`'` 表示不包含, `]` 表示包含
- Hints: 執行 `??pd.cut()`，了解提供其中 `bins` 這個參數的使用方式

作業目標：

- 請同學試著查詢 `pandas.cut` 這個函數還有哪些參數，藉由改動參數以達成目標
- 藉由查詢與改動參數的過程，熟悉查詢函數的方法與理解參數性質，並了解數值的離散化的調整工具

請參考範例程式碼 `Day_012_discretizing`，作業請提交 `Day_012_HW`

[檢視範例](#)

參考資料

課後補充:

連續特徵的離散化：在什麼情況下可以獲得更好的效果(知乎)

這個網頁是個討論串，經由幾個網友的討論與補充，很好地說明了離散化的理由：儲存空間小，計算快，降低異常干擾與過擬合(overfitting)的風險，主要想請同學參考**第1位**的回答，至於其他的討論則請同學參考即可。

網頁連結

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](#)