

Python資料科學程式馬拉松

► Pandas 大型資料處理與效能調校

陪跑專家：Hong



作業

- 作業目標：
 - 了解效能優化方法
- 作業重點：
 - 優化有很多面相可以切入，目前可以先以投影片方法為主

作業題目

1. 在速度較慢的時候，可以先從哪邊開始檢查？
2. 資料過大時應採取什麼方式讓記憶體占用量下降？

1. 在速度較慢的時候，可以先從哪邊開始檢查？

ANS：先檢查有沒有使用到自訂義的函數，如果有可以查詢pandas中是否有支援相關演算法，如果以上都沒辦法改善，那可以從資料讀取方式下手，將要讀取的資料改成pkl檔

2. 資料過大時應採取什麼方式讓記憶體占用量下降？

ANS：將欄位中的型態降級