

百日馬拉松與 Cupoy 問答社區會員啟動

Step1：啟動百日馬拉松會員與 Cupoy 專屬問答社區

1.收到大會寄出會員啟動邀請信(系統將陸續寄出邀請信請耐心等待，感謝!)



2.重新設定密碼

重新設定密碼

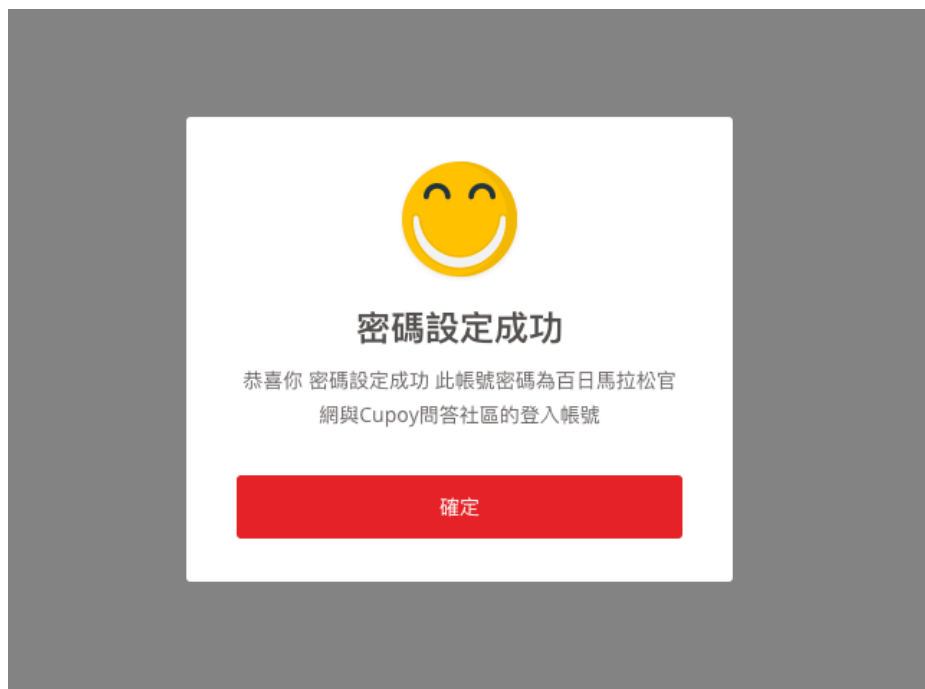
請設定您的登入密碼

請輸入新的密碼

請再輸入一次新密碼

重新設定密碼

3.密碼設定成功



Step2：進入人工智慧知識特助

若知識特助服務為未訂閱，請再重新登入一次。請使用報名 Accupass 時的 email 信箱(參加者的信箱)與新密碼登入。
(完成會員啟動後需等約 3-5 分鐘開啟知識特助權限)

到 Cupoy 首頁點選人工智慧



到知識特助問答社區

[首頁](#)[熱門新聞](#)[知識特助](#)[書籍課程](#)[我的](#)[常見問題](#)

人工智慧

免費體驗

這一年，我們會做你的專屬特助，每日為你速覽、篩選最重要的創業新知，您也可以在問答社區參與討論，我們有專家為您解答，快速掌握新知。

31 人已訂閱

已訂閱

[首頁](#)[最新動態](#)[問答社區](#)[書籍課程](#)

| 每日精選

**進到知識特助
點選問答社區**

[更多](#)

AI城市 | 城市規劃師的新征程：用數據思維喚醒城市的「智商」...



AI應用 | 頻頻被關注的AI，怎樣才能快速落地應用？



AI技術 | 有感 AI 技術，資策會數位轉型研究所加速產業...



自動駕駛 | 騰訊首次發布智慧出行戰略，推出自動駕駛解...



自動駕駛 | 自動駕駛數據到底該歸誰？

如何登入活動官網與提交作業

Step 1 登入官網 <https://ai100-2.cupoy.com/>

1. 使用報名時的電子信箱以及新密碼登入會員



2.提交作業前請先至確認已完成活動前準備軟體安裝與平台帳號申請



3.登入官網，點個人主頁:



Step 2 個人主頁 獲取每日題目

4. 個人主頁介紹_編輯個人檔案

首頁 個人主頁 活動資訊 排名 公告 常見問題 我要提問 登出

答題進度

繼續努力! 已作答 0 題/ 總共釋放 0 題 繼續作答

我的 github 我的 kaggle

應用外文系畢業，認為AI並不是取代人類的科技，而是工具。既然是工具，就要先知道怎麼使用它，他的特性還有他能夠應用在哪些地方。

按此可編輯

編輯個人資料：

1. 顯示頭像
2. 顯示名稱
3. 職稱
4. 關於您
5. 個人 Github 網址
6. 個人 Kaggle 網址

1. 顯示頭像

繼續努力! 已作答 0 題/ 總共釋放 0 題 繼續作答

Floyd 您的職稱 3. 職稱 個人 Github 網址 個人 Kaggle 網址

2. 顯示名稱 關於您 4. 關於您 5. 個人 Github 網址 6. 個人 Kaggle 網址

5. 可觀看與提交已開放題目的作業內容

首頁 個人主頁 活動資訊 排名 公告 常見問題 我要提問 登出

繼續努力! 已作答 0 題/ 總共釋放 1 題 繼續作答

我的 github 我的 kaggle

認為AI並不是取代人類的科技，而是工具。既然是工具，就要先知道怎麼使用它，他的特性還有他能夠應用在哪些地方。

點擊主題

1-資料清理數據前處理 2 1. 資料清理數據前處理 1 未提交
以滾動方式進行資料清理與探索性分析

D1 未提交 資料介紹與評估資料 挑戰是什麼?動手分析前請三思 點擊標題觀看 題目內容

已公布題目 提交作業

D5 未公布 EDA資料分佈 用統計方式描述資料 未公布

未公布題目

6. 題目知識點內容

100 道題目 > D1：資料介紹與評估資料

D1：資料介紹與評估資料

上下頁

Day 1 資料清理數據前處理

資料介紹與評估指標

滾動換上下頁

下載PDF檔案 PDF 下載 全螢幕

7. 作業內容

Sample Code & 作業內容

作業1：

請上 [Kaggle](#), 在 [Competitions](#) 或 [Dataset](#) 中找一組競賽或資料並寫下：

1. 你選的這組資料為何重要
2. 資料從何而來 (tips: 譬如提供者是誰、以什麼方式蒐集)
3. 蒐集而來的資料型態為何
4. 這組資料想解決的問題如何評估

作業2：

想像你經營一個自由載客車隊，你希望能透過數據分析以提升業績，請你思考並描述你如何規劃整體的分析/解決方案：

1. 核心問題為何 (tips：如何定義「提升業績 & 你的假設」)
2. 資料從何而來 (tips：哪些資料可能會對你想問的問題產生影響 & 資料如何蒐集)
3. 蒐集而來的資料型態為何
4. 你要回答的問題，其如何評估 (tips：你的假設如何驗證)

8.檢視範例(有程式碼的作業可檢視範例)

Sample Code & 作業內容

Day 2 作業 (本日不需上傳作業)

- 請下載本次馬拉松建議的 **Kaggle 資料**或準備好自己的資料
- 如採用 Kaggle 資料，請通過 HomeCredit_columns_description.csv，了解各個欄位的意義
- 請點下方檢視範例參考 Day 2 的 ipynb 檔案，了解如何讀取資料並可自行嘗試對資料進行操作

↓ 檢視範例

9.作業檔與範例檔下載，另外提供會需要用到的共用 data 資料夾(按照每個主題下載一次即可，若有更新或異動會再另行通知)

100 道題目 > D1 > 範例下載

範例下載

打包下載

Day_001_example_of_metrics.ipynb

下載

Day_001_HW.ipynb

作業提交檔案
名稱為HW

下載

Data下載 作業有可能會
需要用到的
Data

打包下載(416.8MB)

application_test.csv

25.3MB 下載

application_train.csv

158.4MB 下載

bureau_balance.csv

358.2MB 下載

bureau.csv

162.1MB 下載

credit_card_balance.csv

404.9MB 下載

example.jpg

309.8KB 下載

10. 檢視 ipynb 檔(有程式碼的作業可檢視範例)

檔案詳情

下載檔案

載入套件

```
In [1]: import numpy as np # 用於數學函數的運算
import pandas as pd # 用於讀取以及管理資料
```

讀取資料

首先，我們用 pandas 讀取最主要的資料 application_train.csv (記得到 <https://www.kaggle.com/c/home-credit-default-risk/data> 下載)

Note: ../data/application_train.csv 表示 application_train.csv 與該 .ipynb 的資料夾結構關係如下

```
/data
/application_train.csv
/day_002
/Day_002_first_EDA.ipynb
```



```
In [2]: app_train = pd.read_csv('../data/application_train.csv')
```

印出資料的大小













```
In [3]: print('Training data shape: ', app_train.shape)
```

```
Training data shape: (307511, 122)
```

11. 下載 ipynb 檔執行作業

 jupyter Day_002_first_EDA Last Checkpoint: 17 小時前 (unsaved changes)  Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

           Code 

載入套件

```
In [1]: import numpy as np # 用於數學函數的運算
import pandas as pd # 用於讀取以及管理資料
```

讀取資料

首先，我們用 pandas 讀取最主要的資料 application_train.csv (記得到 <https://www.kaggle.com/c/home-credit-default-risk/data> 下載)

Note: ../data/application_train.csv 表示 application_train.csv 與該 .ipynb 的資料夾結構關係如下

```
/data
/application_train.csv
/day_002
/Day_002_first_EDA.ipynb
```

```
In [2]: app_train = pd.read_csv('../data/application_train.csv')
```

印出資料的大小

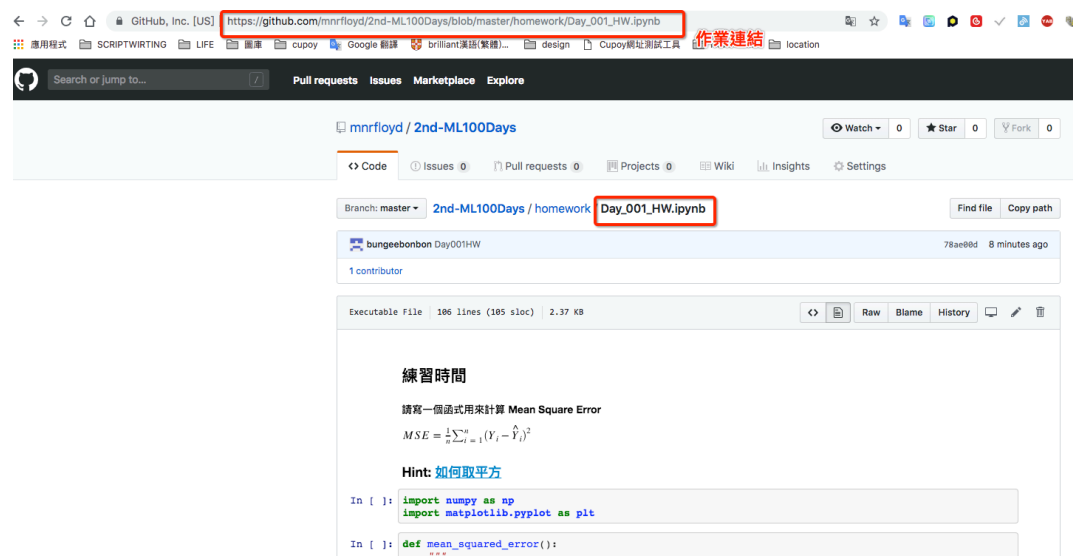
```
In [3]: print('Training data shape: ', app_train.shape)
```

```
Training data shape: (307511, 122)
```

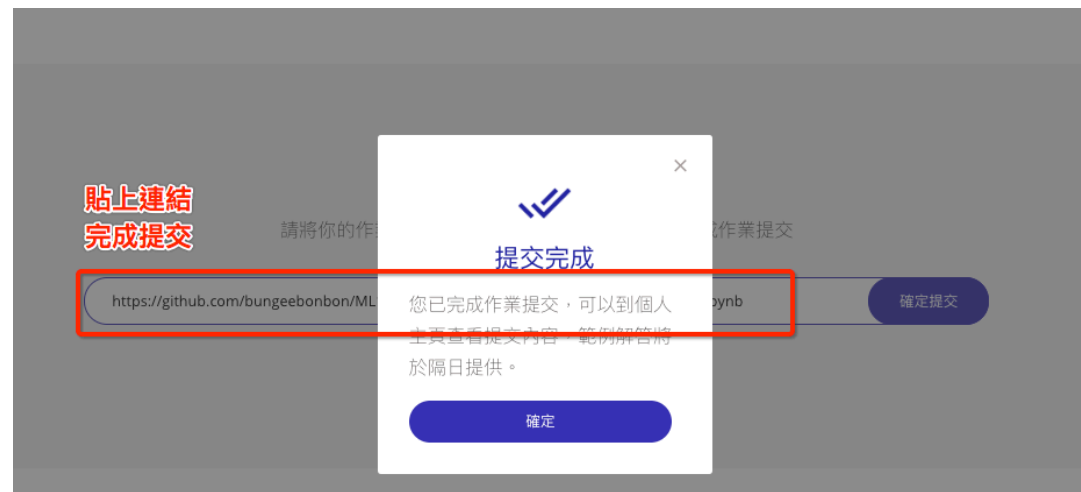
307511 代表的是 row 的數量, 122 代表的是 column 的數量

接下來我們可以用 .head() 這個函數來觀察前 5 row 資料

12. 將作業內容上傳至 Github 上



13 回到官網貼上連結完成提交作業



14. 完成提交狀態/可檢視已提交內容

