

Lab141 : Removing Duplicates

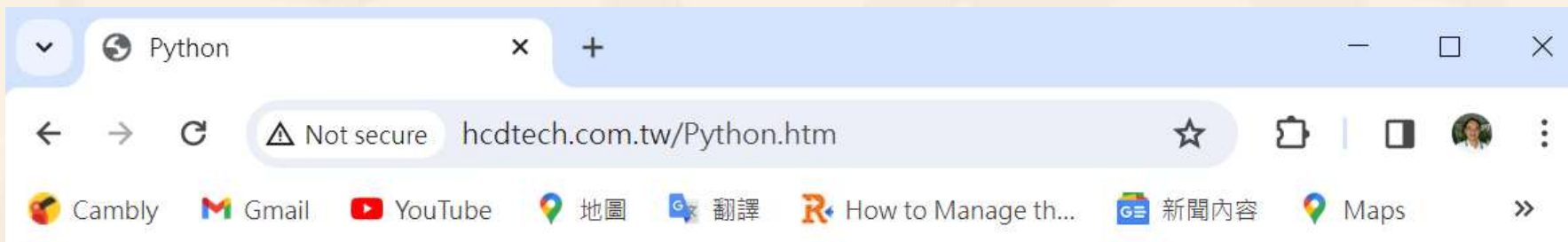
移除重複資料

(別怕! 跟著做你就會!)

盧樹台

shuhtai@uch.edu.tw

請至 www.hcdtech.com.tw 下載教材



<http://www.hcdtech.com.tw/Python.htm>



[\[首頁\]](#) [\[免費研習活動報名須知\]](#) [\[免費自助式教材分享\]](#) [\[Python\]](#) [\[產品簡介\]](#) [\[智慧型遙控器\]](#) [\[汽車震動防盜器\]](#) [\[門窗開啟警報器\]](#) [\[電子密碼鎖\]](#) [\[數位控制電風扇\]](#) [\[房屋電燈中央監控\]](#) [\[洗衣機數控面板\]](#) [\[雙光束雷射防盜器\]](#) [\[火警報知機\]](#)

所有的考卷都可以考100分，是我們自己錯過了！

學習秘訣=發問+練習

考卷發下去，時間到了收回來，如果沒有考到100分，這很正常。重點來了，不會的可以問，問完了練習，準備好了考卷再發下去。第二次還是沒有考到100分，這也很正常。沒關係，再來一次，不會的可以問，問完了練習，準備好了考卷第三次再發下去，.....，考到第N次如果還是沒有考到100分。沒關係，再來，不會的可以問，問完了練習，N+1次、N+2次、.....，你們都很聰明，知道我在說什麼，到最後考卷一定可以考100分！看懂了妳/你就會知道，原來學習的秘訣就是發問和練習！今天開始不懂就問，問完了練習，明年的妳/你肯定不一樣！

學習如何學習！

1

1 2

1 2 3

.....

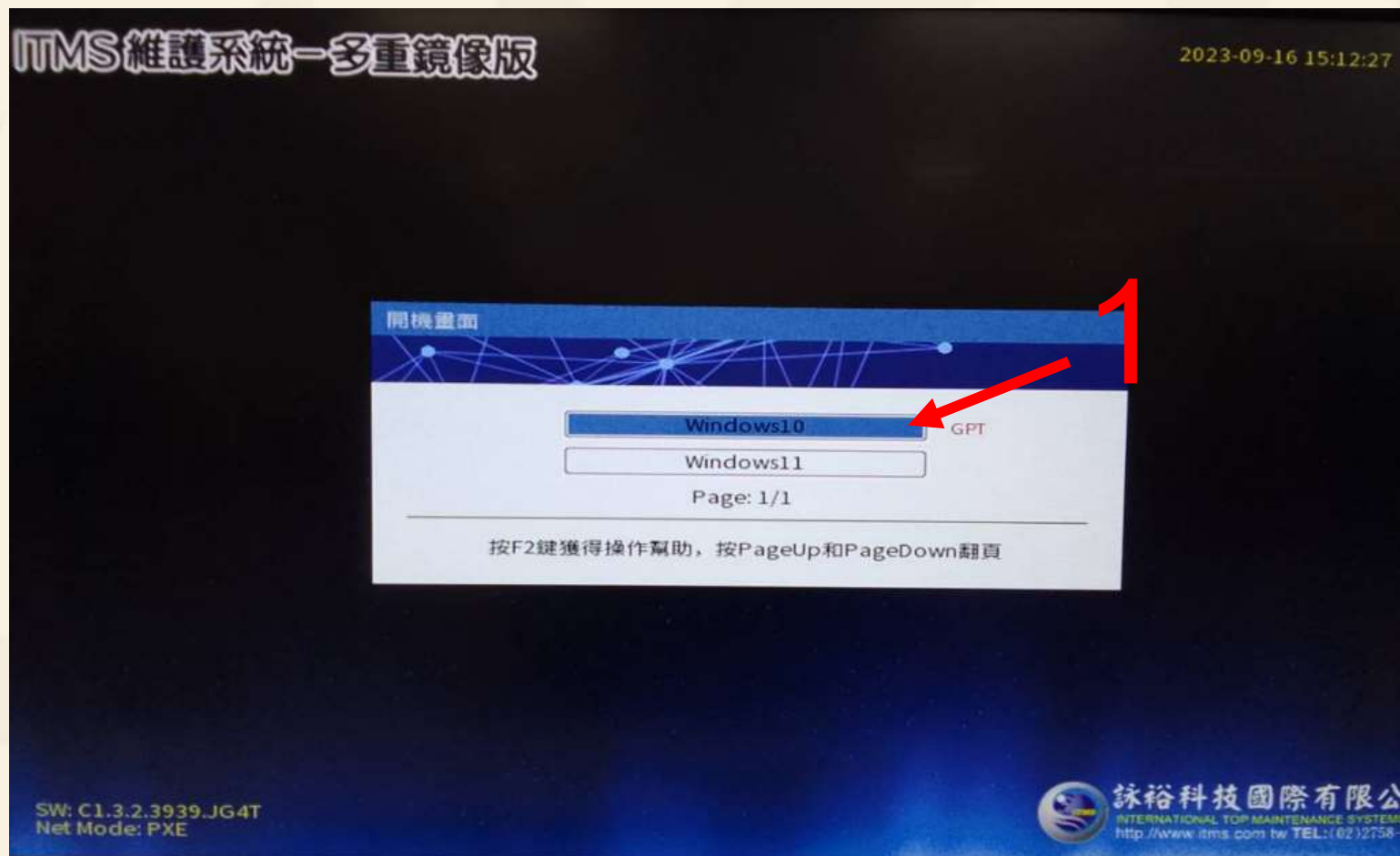
1 2 3 4 5 6 7 8 9 10

金字塔念書法



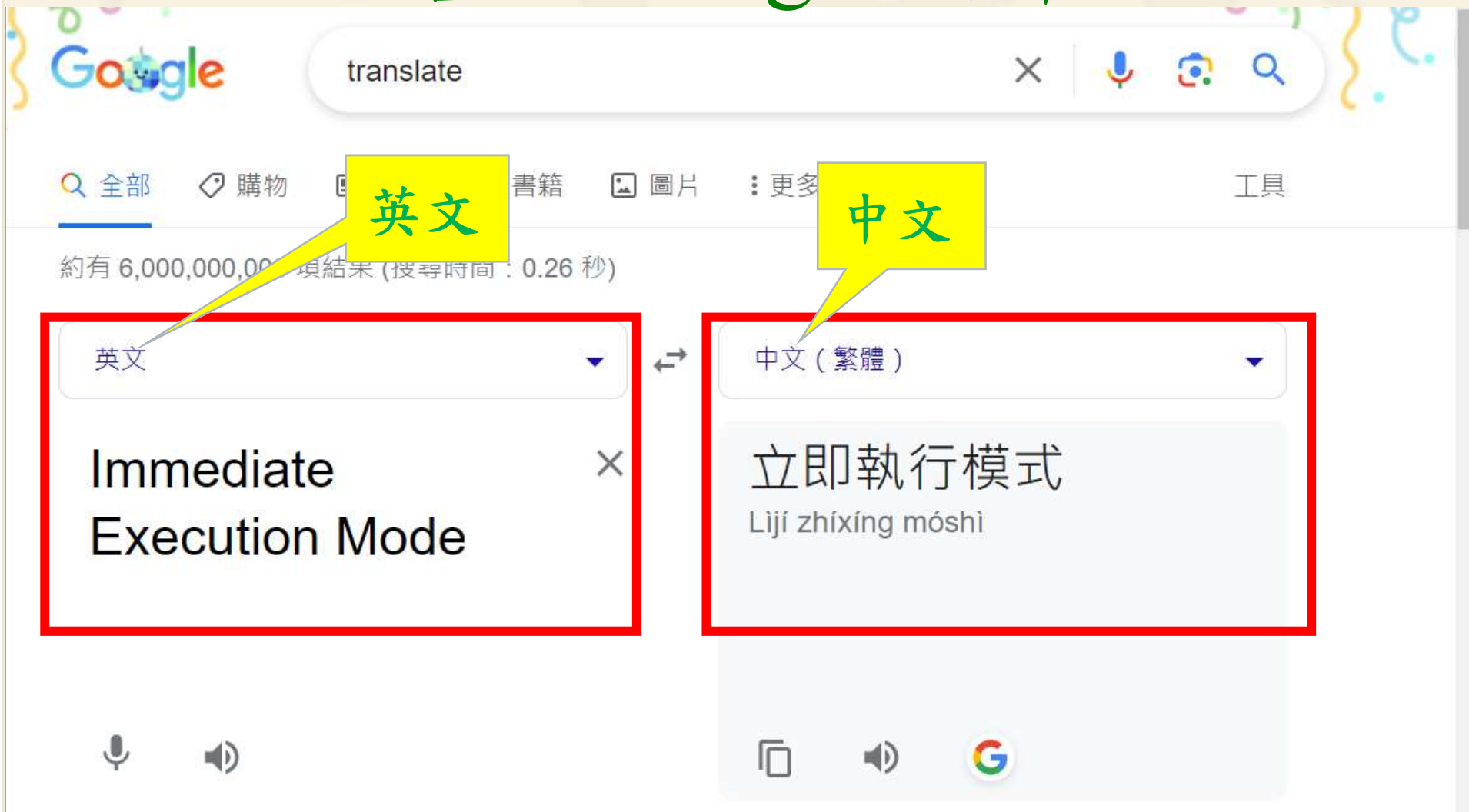
如果一本書有10個章節！先看第1章，在看第2章之前再把第1章看一遍，在看第3章之前再把第1, 2章看一遍，.....，等看到第10章的時候，第1, 2, 3, 4章恐怕已經背起來了！我稱這種念書法為金字塔念書法，今天開始照著做，明年的妳/你肯定不一樣！

請使用 Windows 10



1. 選用 Windows 10.

善用 Google 翻譯



請先開啟網頁閱讀

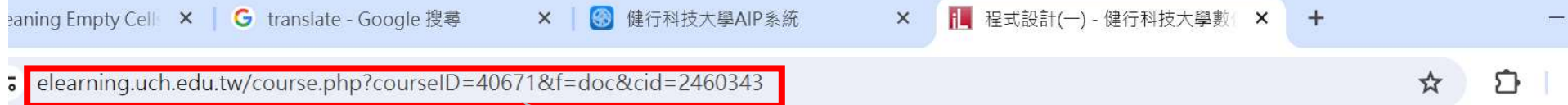
The screenshot shows a web browser window with the URL `w3schools.com/python/pandas/pandas_cleaning_duplicates.asp` highlighted in the address bar. A yellow callout box with the text "請用善用 Google 翻譯 讀懂 網頁 內容" (Please use Google Translate to understand the website content) points to the URL. The page content includes a navigation menu on the left, a main heading "Pandas - Removing Duplicates", and a table of data with columns: Duration, Date, Pulse, Maxpulse, and Calories.

	Duration	Date	Pulse	Maxpulse	Calories
0	60	'2020/12/01'	110	130	409.1
1	60	'2020/12/02'	117	145	479.0
2	60	'2020/12/03'	103	135	340.0
3	45	'2020/12/04'	109	175	282.4
4	45	'2020/12/05'	117	148	406.0
5	60	'2020/12/06'	102	127	300.0

請用善用 Google 翻譯 讀懂 網頁 內容



本練習請先下載CSV檔



1. 請至iLMS網頁

2. 下載Lab136data.csv後
將檔案複製到Win10(C)>
使用者>User資料夾

課程: 程式設計(一) (ETP002A)

瀏覽模式 [切換]
您現在的瀏覽模式為 [老師]

- 課程功能 [管理]
- 課程活動(公告)
- 上課教材 (7)
- 課堂整理
- 課程說明
- 課程行事曆
- 討論區
- 小組專區
- 隨堂筆記 (共享的筆記)
- 作業
- 問卷
- 線上測驗
- 出缺勤 (統計)
- 成員 (15)

位置: 程式設計(一) > 上課教材 > 教材

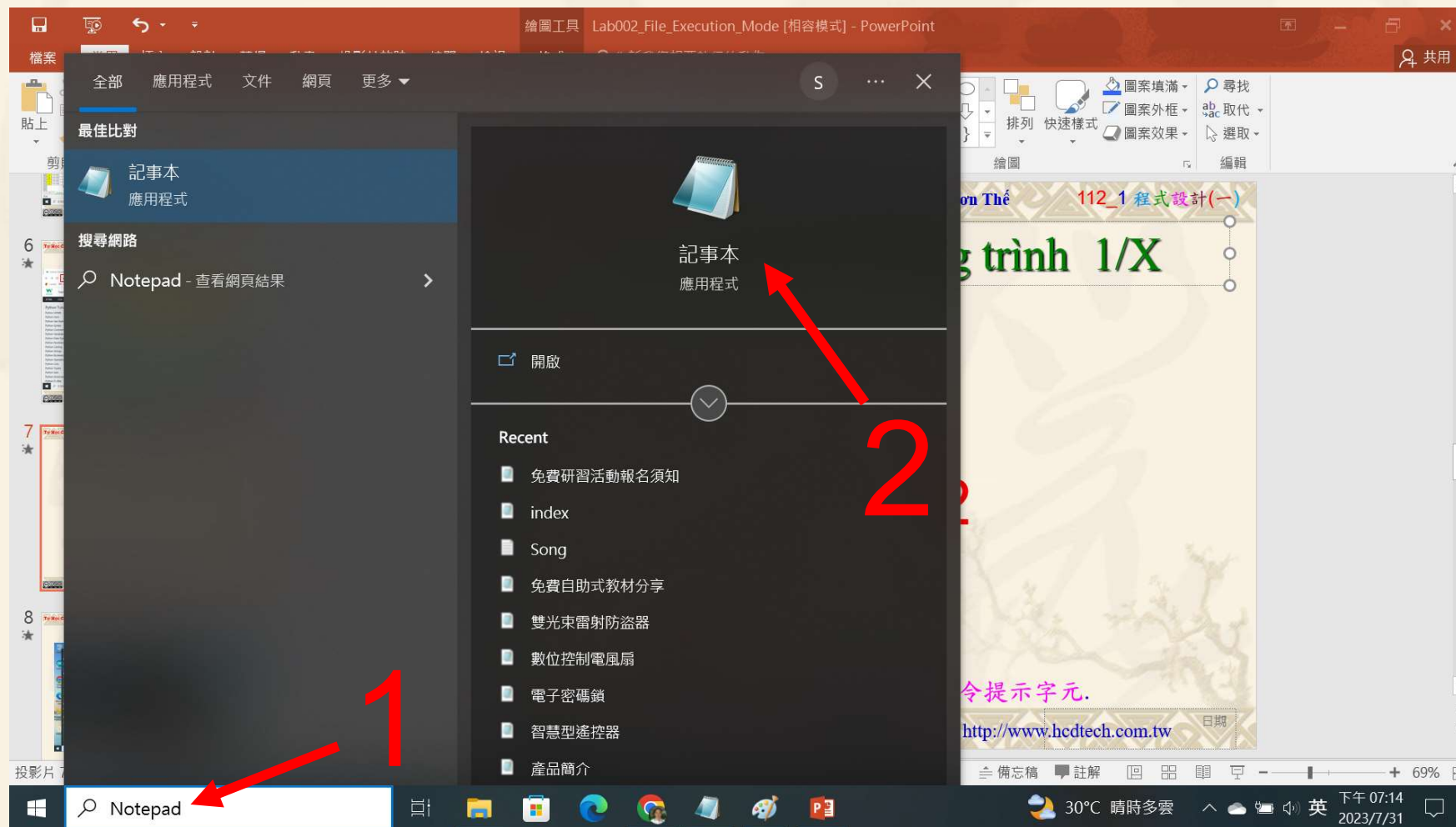
Lab126~Lab150
by 盧樹台, 2023-11-30 20:51, 人氣(9)

- 附件 [隱藏]
- 1. Lab126_NumPy GCD.pdf (1.4 MB)
- 2. Lab127_Trigonometric Functions.pdf (1.4 MB)
- 3. Lab128_Trigonometric Functions.pdf (1.4 MB)
- 4. Lab129_Hyperbolic Functions.pdf (1.4 MB)
- 5. Lab130_Set Operations.pdf (1.7 MB)
- 6. Lab131_Pandas Getting Started.pdf (1.4 MB)
- 7. Lab132_Pandas DataFrames.pdf (1.4 MB)
- 8. Lab133_Pandas Read CSV.pdf (1.4 MB)
- 9. Lab134_Pandas Read JSON.pdf (1.4 MB)
- 10. Lab135_Analyzing DataFrames.pdf (1.6 MB)
- 11. Lab136data.csv (1 KB)
- 12. Lab136_Cleaning Empty Cells.pdf (1.6 MB)

發表討論



建立程式文件 1/4



1. 鍵盤輸入Notepad. 2. 用滑鼠點選記事本.

建立程式文件 2/4

*未命名 - 記事本

檔案(F) 編輯(E) 格式(O) 檢視(V) 說明

```
print("P11211XXX practices Lab141.")
```

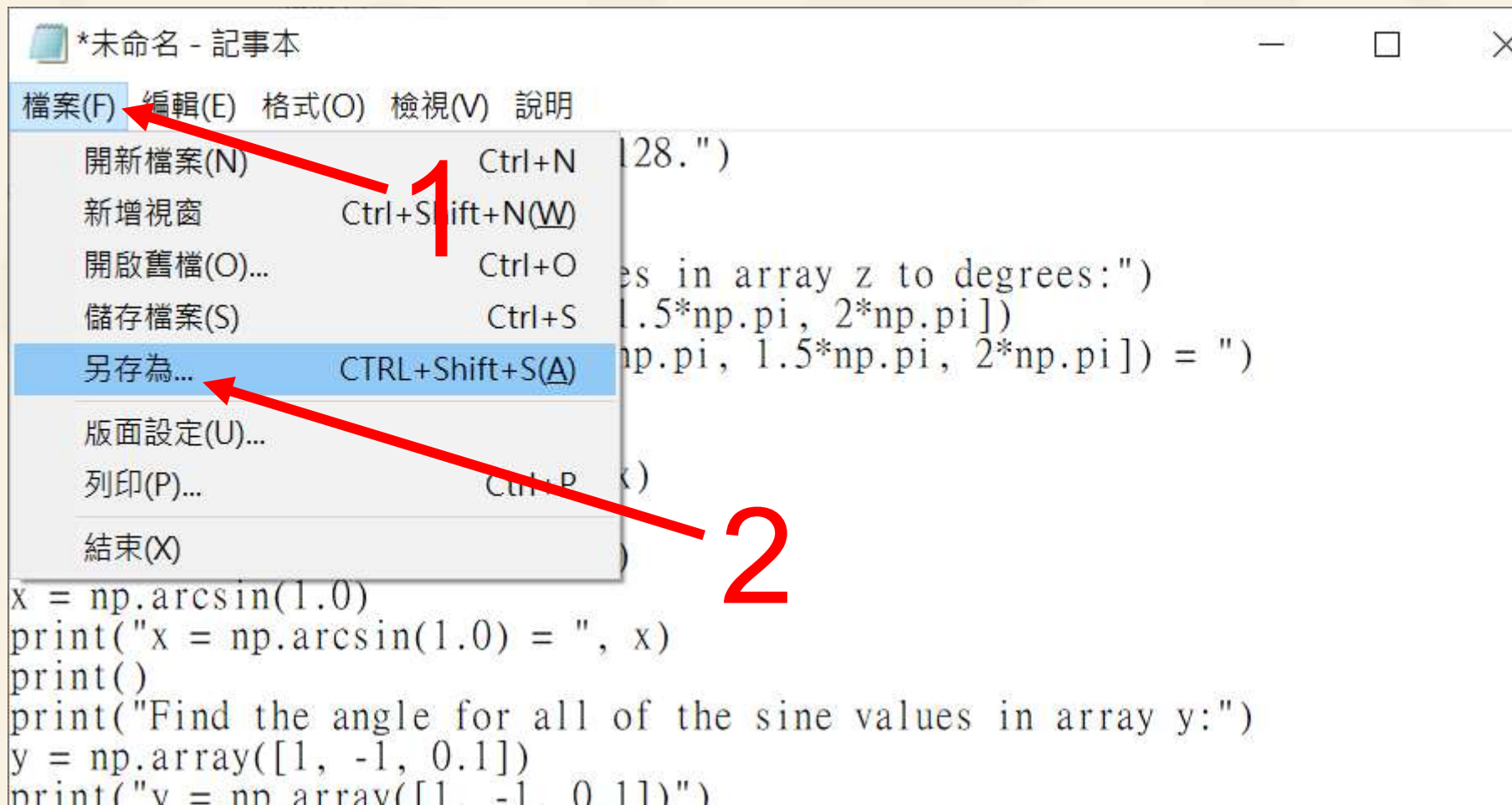
將P11211XXX修改為您的學號

```
import pandas as pd
print("Load Lab136data.csv into a DataFrame:")
MyDataFrame = pd.read_csv('Lab136data.csv')
print(MyDataFrame.to_string())
print("The data set contains some duplicate rows (Check out row 11 and row 12).")
x = input("Please check duplicate rows in row 11 and row 12 and press the 'Enter' key on your keyboard.")
print()
print("Returns True for every row that is a duplicate, otherwise False:")
print(MyDataFrame.duplicated())
x = input("Please check row 12 returns True and press the 'Enter' key on your keyboard.")
print()
print("Remove all duplicates:")
MyDataFrame.drop_duplicates(inplace = True)
print(MyDataFrame.to_string())
x = input("Please check all duplicates are removed and press the 'Enter' key on your keyboard.")
print("P11211XXX practices Lab141.")
```

將P11211XXX修改為您的學號

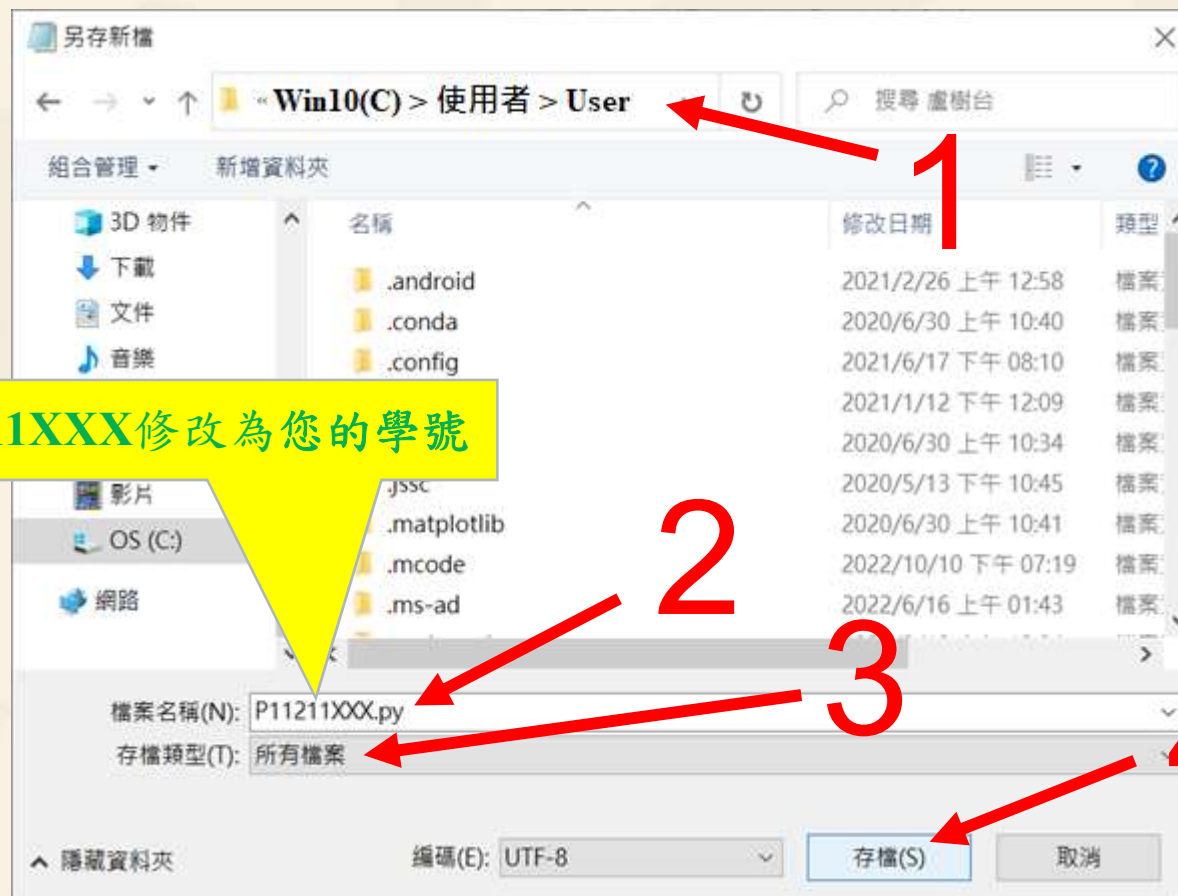
1. 用鍵盤輸入程式代碼.

建立程式文件 3/4



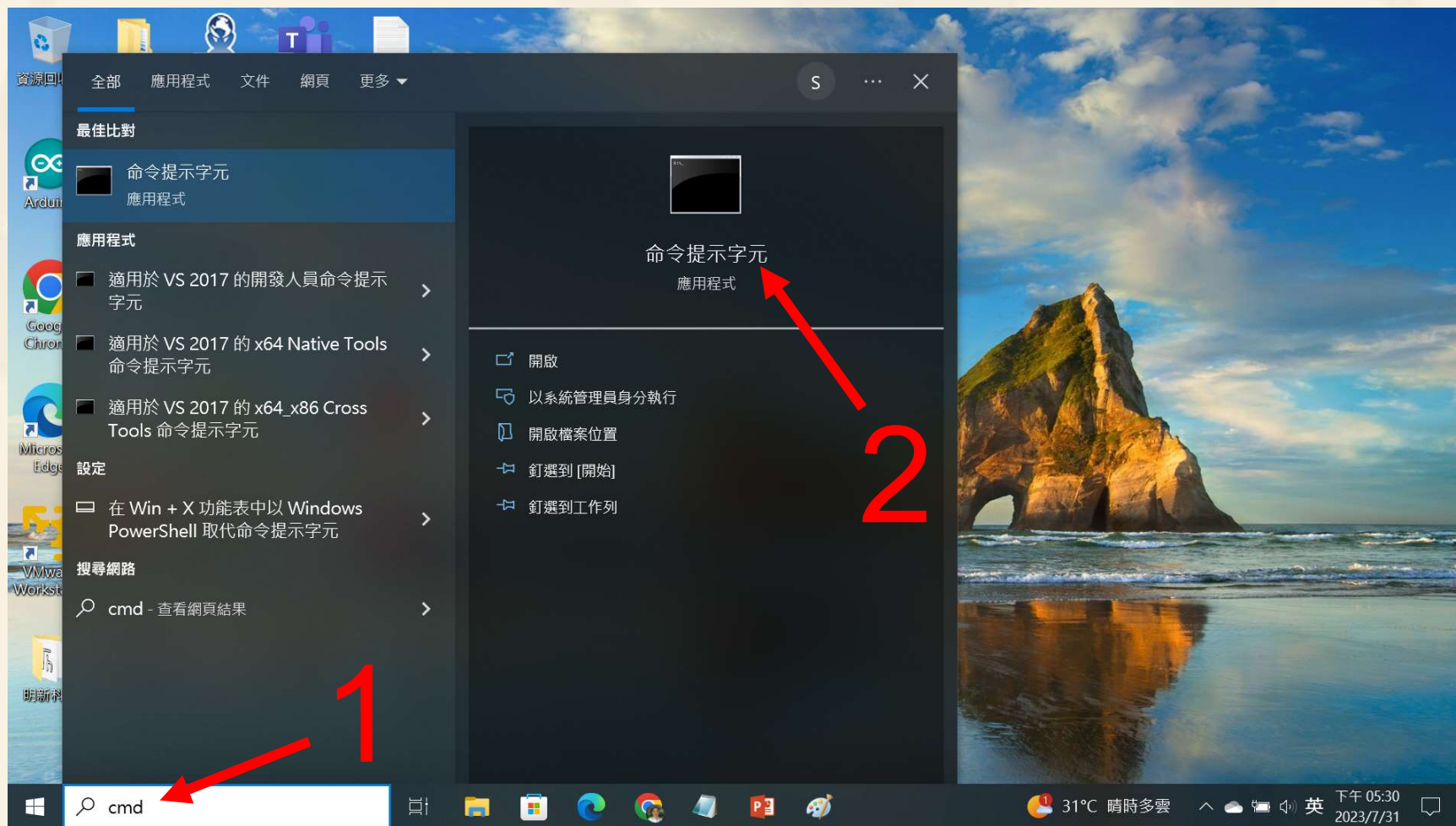
1. 用滑鼠點選檔案. 2. 用滑鼠點選另存為....

建立程式文件 4/4



1. 資料夾 = C:\使用者>User>.
2. 檔案名稱 = P11211XXX.py .
3. 存檔類型(T) = 所有檔案.
4. 用滑鼠點選存檔.

檔案執行模式 1/3



1. 鍵盤輸入cmd.
2. 用滑鼠點選命令提示字元.

檔案執行模式 2/3

```
C:\> 命令提示字元  
Microsoft Windows [版本 10.0.19045.3803]  
(c) Microsoft Corporation. 著作權所有，並保留一切權利。  
C:\Users\盧樹台>pip install pandas
```



- 1. 用鍵盤輸入pip install pandas.
- 2. 按一下Enter.

檔案執行模式 3/3

```
C:\> 命令提示字元

Microsoft Windows [版本 10.0.19045.3803]
(c) Microsoft Corporation. 著作權所有，並保留一切權利。

C:\Users\盧樹台>pip install pandas
Requirement already satisfied: pandas in c:\python39\lib\site-packages (2.1.3)
Requirement already satisfied: six>=1.5 in c:\python39\lib\site-packages (from
WARNING: Ignoring invalid distribution -ip (c:\python39\lib\site-packages)
WARNING: Ignoring invalid distribution - (c:\python39\lib\site-packages)

C:\Users\盧樹台>Python P11211XXX.py
```

將P11211XXX修改為您的學號



1. 用鍵盤輸入Python P11211XXX.py .
2. 按一下Enter.

Verification Criteria of Lab141

(Lab141的驗收規範) 移除重複資料

**P11211XXX 必需
更換為您的學號**

Ask the teacher to give you points after completing the illustrated results.

(完成右圖指定成果後請教師在您的座位驗收並讓您簽名加分)

命令提示字元

```

4      45  '2020/12/05'  117  148  406.0
5      60  '2020/12/06'  102  127  300.0
6      60  '2020/12/07'  110  136  374.0
7      450 '2020/12/08'  104  134  253.3
8      30  '2020/12/09'  109  133  195.1
9      60  '2020/12/10'  98   124  269.0
10     60  '2020/12/11'  103  147  329.3
11     60  '2020/12/12'  100  120  250.7
13     60  '2020/12/13'  106  128  345.3
14     60  '2020/12/14'  104  132  379.3
15     60  '2020/12/15'  98   123  275.0
16     60  '2020/12/16'  98   120  215.2
17     60  '2020/12/17'  100  120  300.0
18     45  '2020/12/18'  90   112   NaN
19     60  '2020/12/19'  103  123  323.0
20     45  '2020/12/20'  97   125  243.0
21     60  '2020/12/21'  108  131  364.2
22     45   NaN        100  119  282.0
23     60  '2020/12/23'  130  101  300.0
24     45  '2020/12/24'  105  132  246.0
25     60  '2020/12/25'  102  126  334.5
26     60   20201226    100  120  250.0
27     60  '2020/12/27'  92   118  241.0
28     60  '2020/12/28'  103  132   NaN
29     60  '2020/12/29'  100  132  280.0
30     60  '2020/12/30'  102  129  380.3
31     60  '2020/12/31'  92   115  243.0
Please check all duplicates are removed and press the 'Enter' key on your keyboard.
P11211XXX practices Lab141.
C:\Users\盧樹台>
                
```

P11211XXX - 記事本

```

檔案(F) 編輯(E) 格式(O) 檢視(V) 說明
print("P11211XXX practices Lab141.")

import pandas as pd
print("Load Lab136data.csv into a DataFrame:")
MyDataFrame = pd.read_csv('Lab136data.csv')
print(MyDataFrame.to_string())
print("The data set contains some duplicate rows (Che
x = input("Please check duplicate rows in row 11 and
print()
print("Returns True for every row that is a duplicate
print(MyDataFrame.duplicated())
x = input("Please check row 12 returns True and press
print()
print("Remove all duplicates:")
MyDataFrame.drop_duplicates(inplace = True)
print(MyDataFrame.to_string())
x = input("Please check all duplicates are removed an
print("P11211XXX practices Lab141.")
                
```

每一個學生都要做Lab141至少一次!

養成良好的工作態度

- 離開實驗室時請整理自己的工作座位，為自己的工作態度加分：
 - (1)滑鼠鍵盤歸位 (2)電腦關機 (3)螢幕關閉電源 (4)椅背靠妥 (5)個人責任區(工作座位及週邊範圍)應整潔，不遺留垃圾紙屑等。

