Google Developer Student Clubs

Data Processing with Python

GDSC NTNU 2023/11/13

Hugo Wang
@whyhugo

Google Developer Student Clubs

whoami

Hugo Wang

@whyhugo

- 師大資工 大一
- GDSC NTNU Tech Core Team Member
- SITCON 2024 議程組副組長
- 教育大數據微學程 通識 TA

Outline

You will learn…

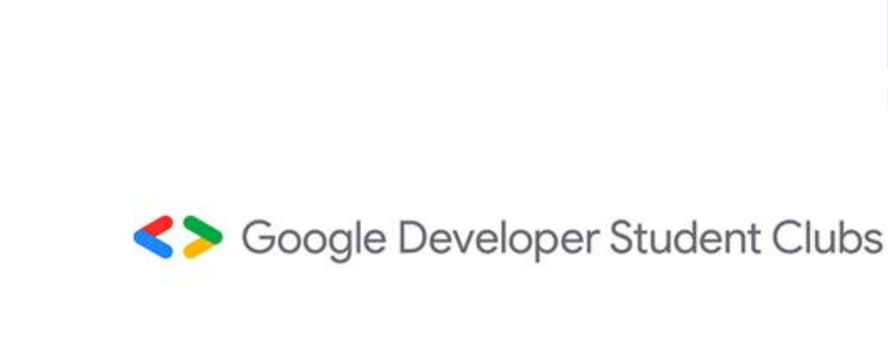
. 資料處理開發環境
. 資料哪裡來
. Numpy & Pandas 實作
. Matplotlib 實作

今日 GitHub：whyhugo/GDSC-NTNU-handouts

資料處理開發環境

#intro

- 區塊化 (cell) 編程
  - 容易 debug
  - 分段處理、觀察
- Preview
- Markdown 筆記
- 保存結果共享



資料哪裡來？

#intro

爬蟲

Open data

By gov

GitHub

Cmt.

Company/
Customer

Kaggle

ML Competition

Numpy

#data

- 資料科學和機器學習常用的資料結構
- 裝載相同類型資料的多維陣列
- 較 List 高效且擁有多項功能，使用上須注意與 List 的差異
- shape：資料維度, ex. (2,3) 代表 2x3 的矩陣
- dtype：資料型態, 陣列中元素的型態

Pandas

#data

- Numpy 是以 Array 形式呈現
- Pandas 是以 Numpy 為基礎，提供三種資料結構：
  - Series (一維)、DataFrame (二維)、Panel (三維)
- Series 可看成可調整 index 的 Array，使用上須注意位置和 index 的差別
- DataFrame 可看成可調整 index 和 column 的多個 Series

Workshop

#try

Education	Traffic	Environment	Travel
大專院校男女比資料集樞紐分析	YouBike 熱門站點與建置成本關係	翡翠水庫卡爾森優養指數與氣象數據	貓空纜車運量與氣象數據

Google Developer Student Clubs

THANK YOU

:D

匿名回饋表單

今日課程資料