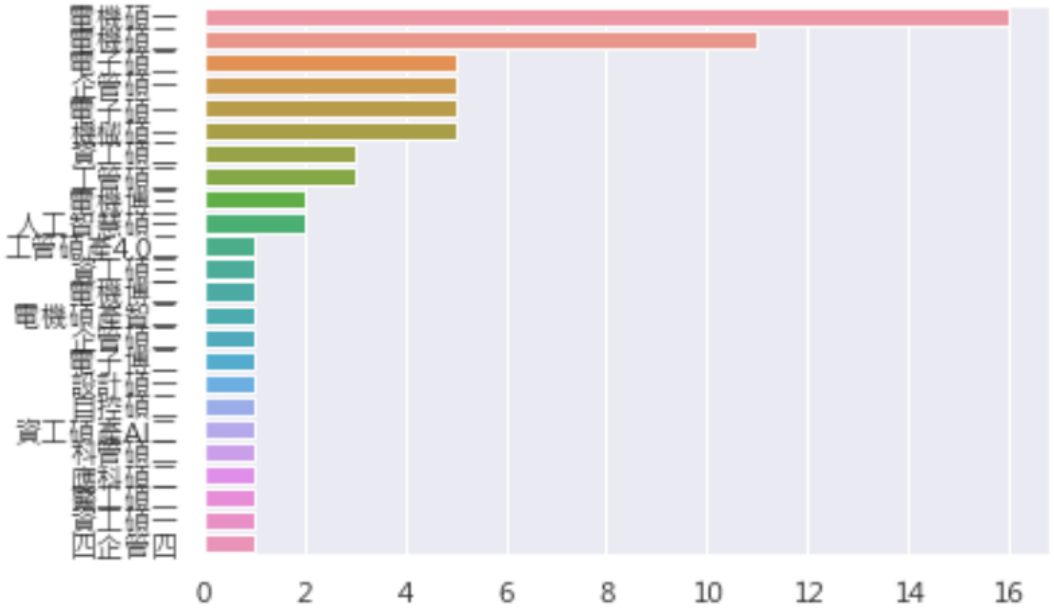
Script of the data science
2022/9/6, 2021/9/24

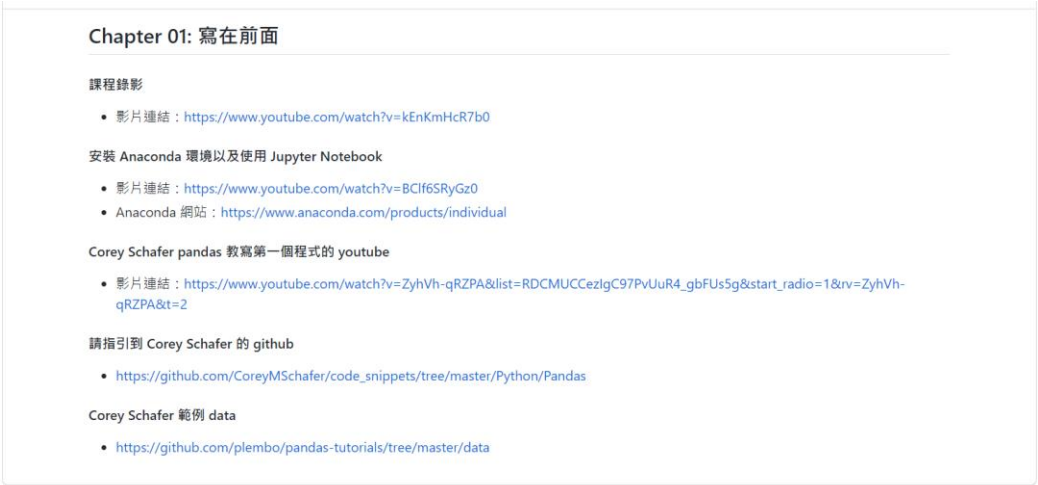
1. Disclaimer: This is NOT a big data course, but a data science course.
2. However, we DO cover deep learning-based approaches, which require A LOT of data, for e-Commerce applications through term projects.

3. Email notification of the Google meet room to all registered students
4. Data analysis of registered students



5. You have to be Python experienced to feel most productive in taking this course.

- 課程 github: https://github.com/NTUST-ICLab-Course/Data-Analysis_v1111
  去年的 github https://github.com/NTUST-ICLab-Course/Data-Analysis_v1101



Chapter 01: 寫在前面

課程錄影

- 影片連結：https://www.youtube.com/watch?v=kEnKmHcR7b0

安裝 Anaconda 環境以及使用 Jupyter Notebook

- 影片連結：https://www.youtube.com/watch?v=BCIf6SRyGz0
- Anaconda 網站：https://www.anaconda.com/products/individual

Corey Schafer pandas 教寫第一個程式的 youtube

- 影片連結：https://www.youtube.com/watch?v=ZyhVh-qRZPA&list=RDCMUCCezIgC97PvUuR4_gbFUs5g&start_radio=1&rv=ZyhVh-qRZPA&t=2

請指引到 Corey Schafer 的 github

- https://github.com/CoreyMSchafer/code_snippets/tree/master/Python/Pandas

Corey Schafer 範例 data

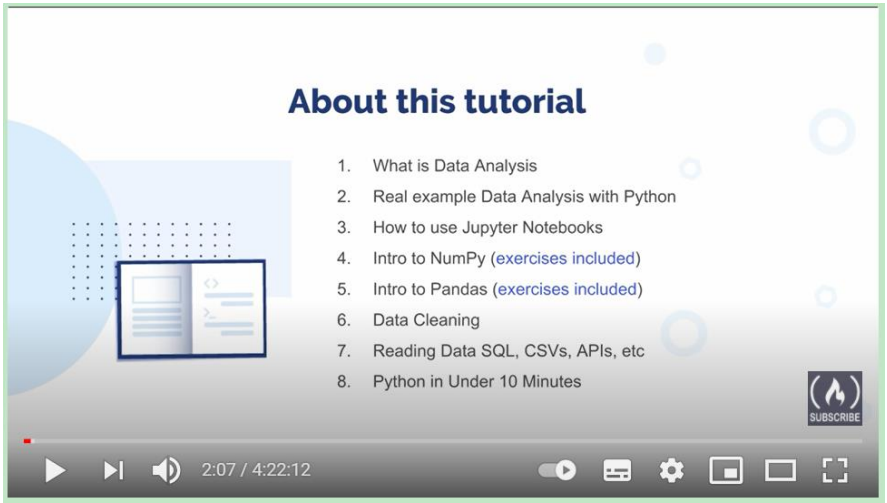
- https://github.com/plembo/pandas-tutorials/tree/master/data

- 評分機制：
  - 資料標註練習 2 次：10 * 2 = 20%
  - (兩人隨機配對一組)習題 5 次：8 * 5 = 40%
  - (三人一組)分組專題，期中指定題與期末自選題各一次：20*2 = 40%

- 習題：五題
  - 網路爬文技術 Beautifulsoup，Selenium (旭清)；
  - 臺灣證券交易所每日股票買賣超主力分析 (旭清)；
  - 不同電子商務公司之搜尋與價格比較 (竣崴)；
  - 電影之搜尋與推薦系統 (可東、竣崴)；
  - 總統大選選舉捐款金源分析 (威霆)；

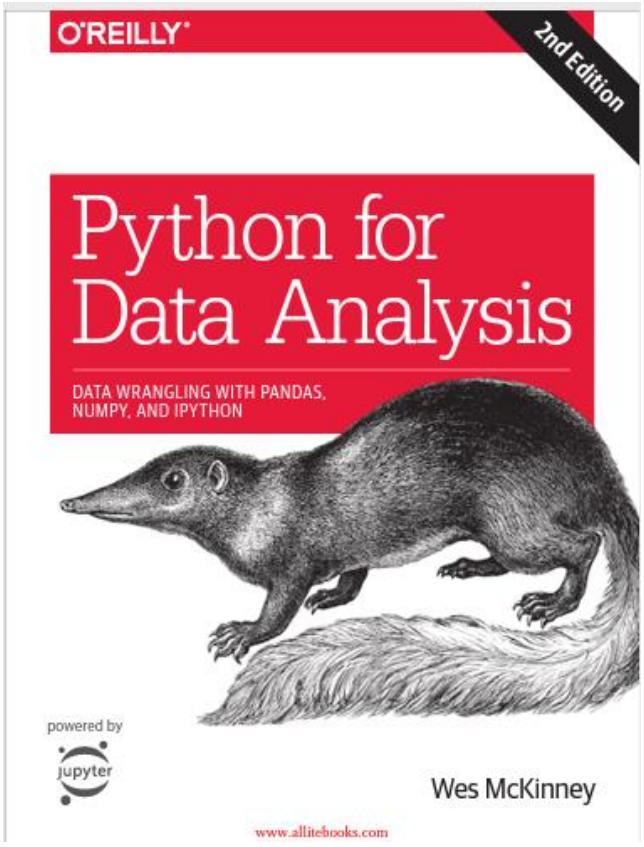
- 學期專題：期中指定與期末自選
  - 期中指定題為電子商務的比較：
    (1) 搜尋結果→ 用 Round1 的測試集作評測；
    (2) 比價：同商品的價格。電子商務商家包括：京東、Momo，阿里巴巴，最後是 PChome，等。
  - 期末自選：自己決定，或是助教指導的
    - 3 組作「電子商務之搜尋模型訓練與驗證集」 (竣崴)；
    - 3 組作「特定分類商品之標籤辨識技術」NER (威霆)；
    - 3 組作「客服聊天機器人」(旭清)。

6. Several good online tutorials on general modules of data science, like: numPy, pandas, matplotlib, and seaborn, etc.
   https://www.youtube.com/watch?v=r-uOLxNrNk8&t=113s

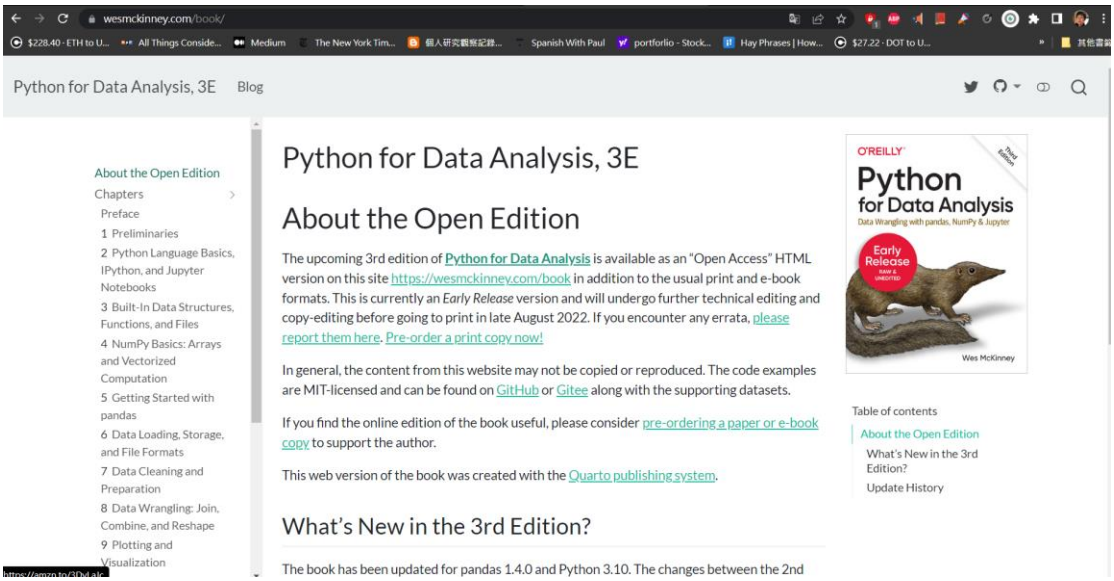


Data Analysis with Python - Full Course for Beginners (Numpy, Pandas, Matplotlib, Seaborn)

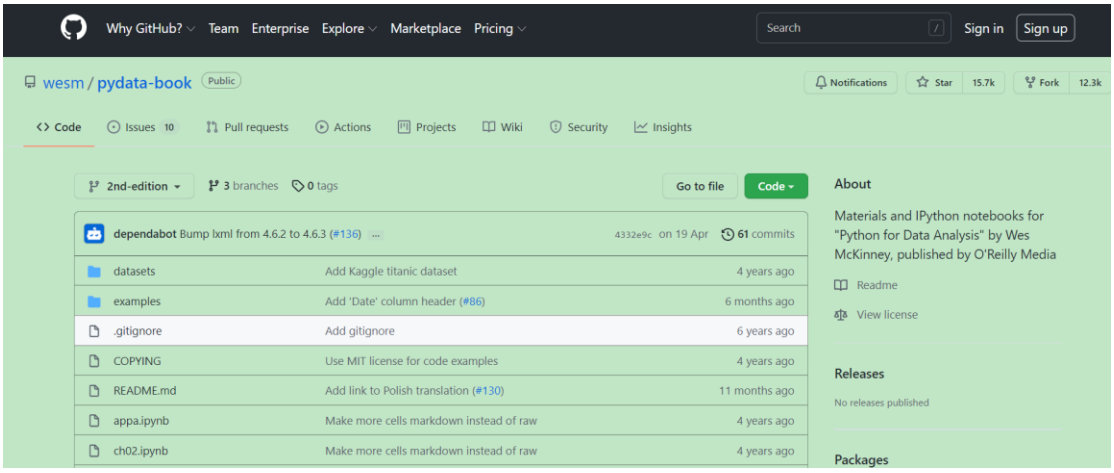
7. Textbook pdf, somehow, is available online:
   https://www.programmer-books.com/wp-content/uploads/2019/04/Python-for-Data-Analysis-2nd-Edition.pdf



8. Actually, the new third edition is (?) just out! And an “Open Access” HTML version is available here: https://wesmckinney.com/book/

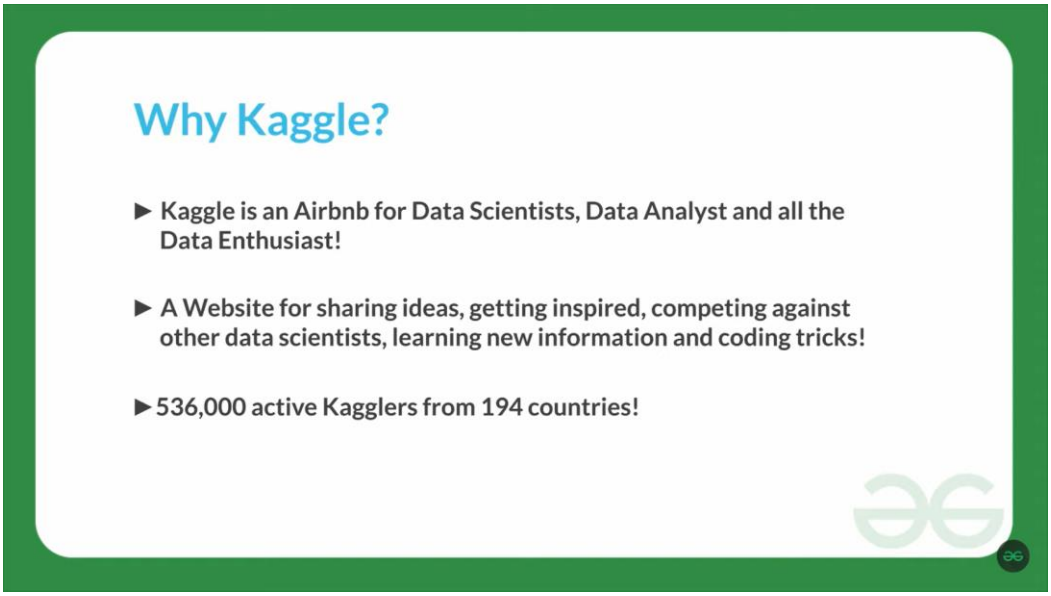


9. github of the Python codes in the textbook by the author: (second edition) https://github.com/wesm/pydata-book

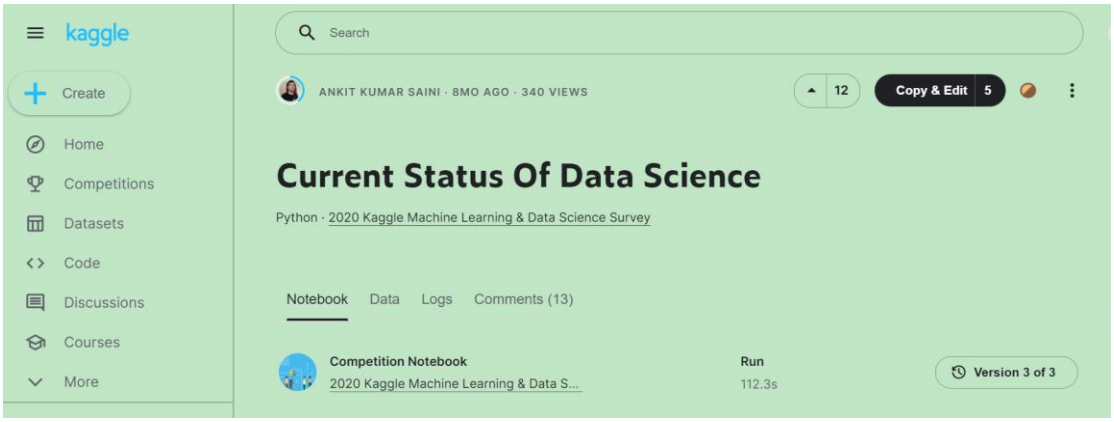


github of the Python codes in the textbook by the author: (third edition) https://github.com/wesm/pydata-book/tree/3rd-edition

10. Datasets exploration from Kaggle for individual project proposal and term project. https://www.youtube.com/watch?v=u9MIwoFWXVg&t=79s

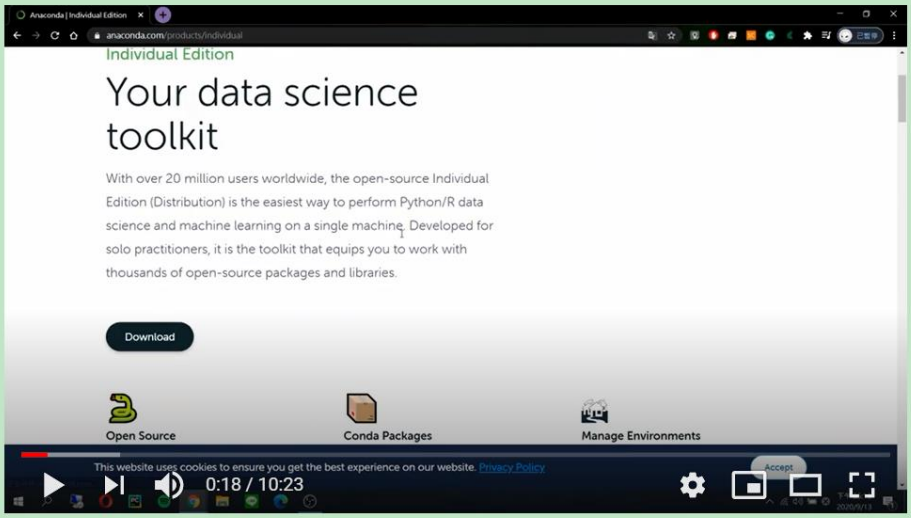


11. Current status of data science:
    https://www.kaggle.com/ankitkumarsaini/current-status-of-data-science



12. Install anaconda + Jupyter

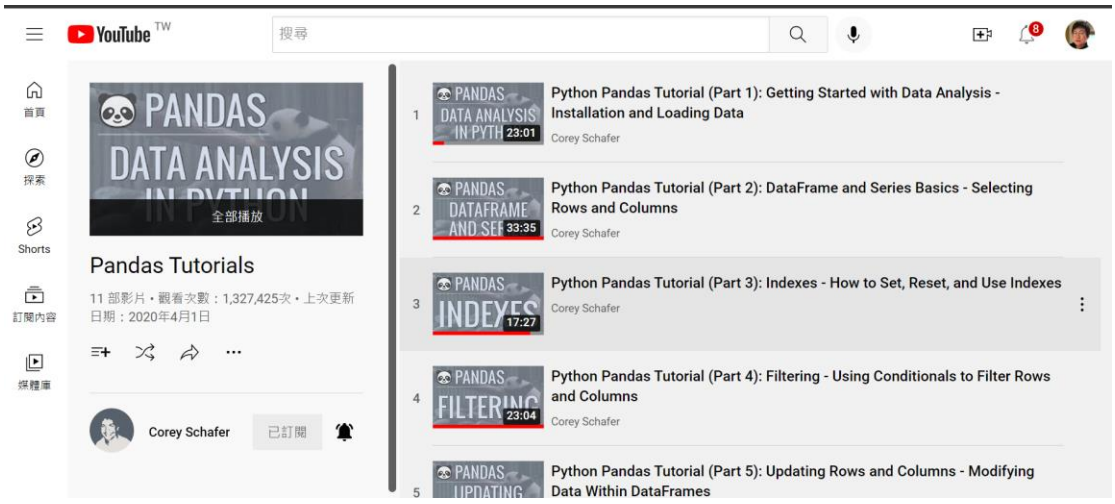
https://www.youtube.com/watch?v=BClf6SRyGz0&t=15s



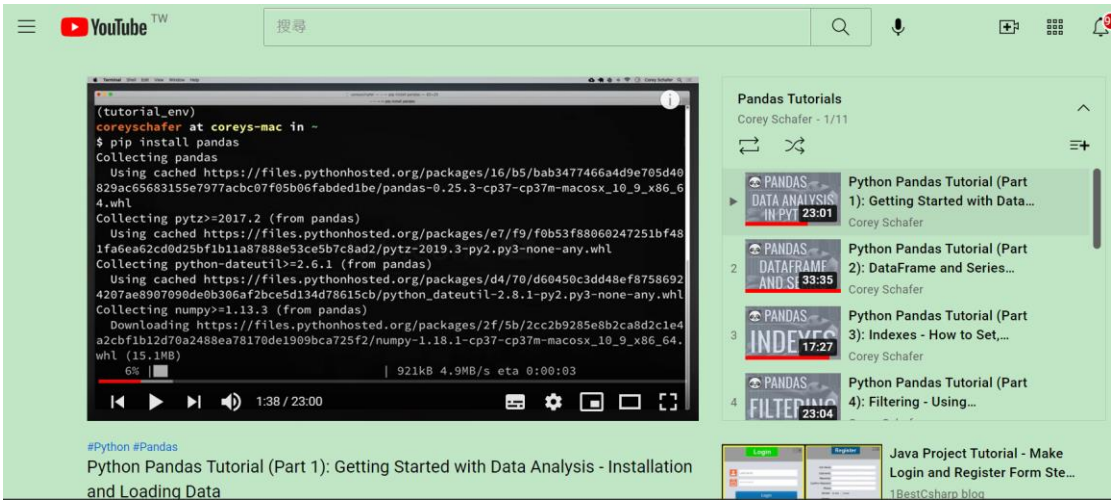
安裝 Anaconda 環境以及使用 Jupyter Notebook

13. Try out Corey Schafer's Pandas tutorial series

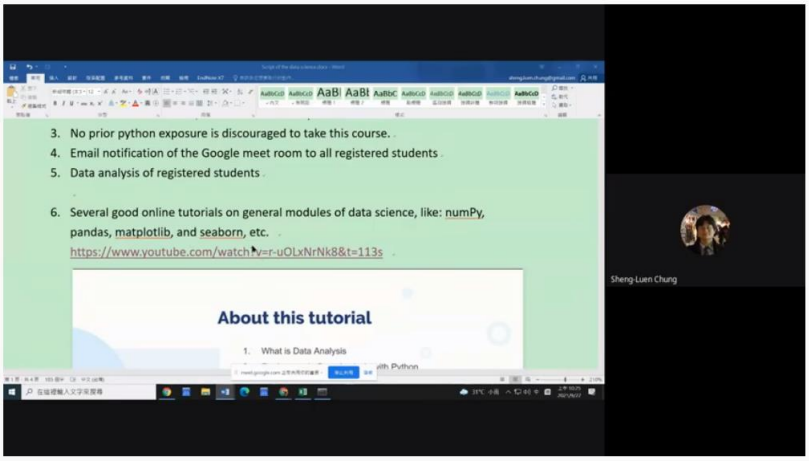
https://www.youtube.com/playlist?list=PL-osiE80TeTsWmV9i9c58mdDCSskIFdDS



https://www.youtube.com/watch?v=ZyhVh-qRZPA&list=PL-osiE80TeTsWmV9i9c58mdDCSskIFdDS



First (Introductory) week of this course:
https://www.youtube.com/watch?v=e544uSMYP-c



pfa jbhe zco 2021 09 26 at 19 25 GMT 7

Takeaway:

1. Install anaconda + jupyter
   https://www.youtube.com/watch?v=BClf6SRyGz0&t=15s
2. Finish reading Chapter 2 of the textbook (2nd Ed. or 3rd Ed.)
3. Try out ch02.ipynb (TA will announce when and where it is ready)
4. Try out Corey Schafer's Python Pandas Tutorial (part 1)
   https://www.youtube.com/playlist?list=PL-osiE80TeTsWmV9i9c58mdDCSskIFdDS
5. Find two other of your project team members.
6. I will not see you next week. Please check out the recorded video, announced later in Moodle.
   https://github.com/NTUST-ICLab-Course/Data-Analysis_v1111