

A man in a brown jacket and blue jeans is longboarding down a winding road. The background features a range of mountains under a bright sky.

# 科系、職務與待遇： 大數據時代下的社會學與勞動市場

- 游騰林
- 2020-11-20



# 游騰林 數據分析師

---



東華大學 社會學研究所



國泰世華 數據部



<https://github.com/TLYu0419>



# 提問方式介紹

別懷疑趕快去下載 App ~

slido

Join at  
**Slido.com**  
**#44701**

# 課程大綱

迷路的時候就看看右上角吧！

關於社會學

數據分析

傑華老師的煩惱

總結

關於社會學

數據分析

傑華老師的煩惱

總結

- |           |           |               |        |
|-----------|-----------|---------------|--------|
| • 社會學教什麼? | • 分析流程    | • 院長突然約傑華吃午餐？ | • 紿系上  |
| • 畢業之後呢   | • 反面應用案例  | • Python 簡介   | • 紿學弟妹 |
| • 職涯規劃    | • 大數據與社會學 | • 網路爬蟲實做      | • 學習資源 |
|           |           | • 資料分析與視覺化    |        |

A man in a maroon hoodie and tan pants is longboarding down a winding asphalt road. He is leaning into the turn, looking towards the camera. The background features rolling green hills and mountains under a bright, slightly cloudy sky.

# 關於社會學

- 社會學教什麼？
- 畢業之後呢？
- 職涯規劃

# 到底社會學在教什麼>"<

# 社會系課程規劃理念

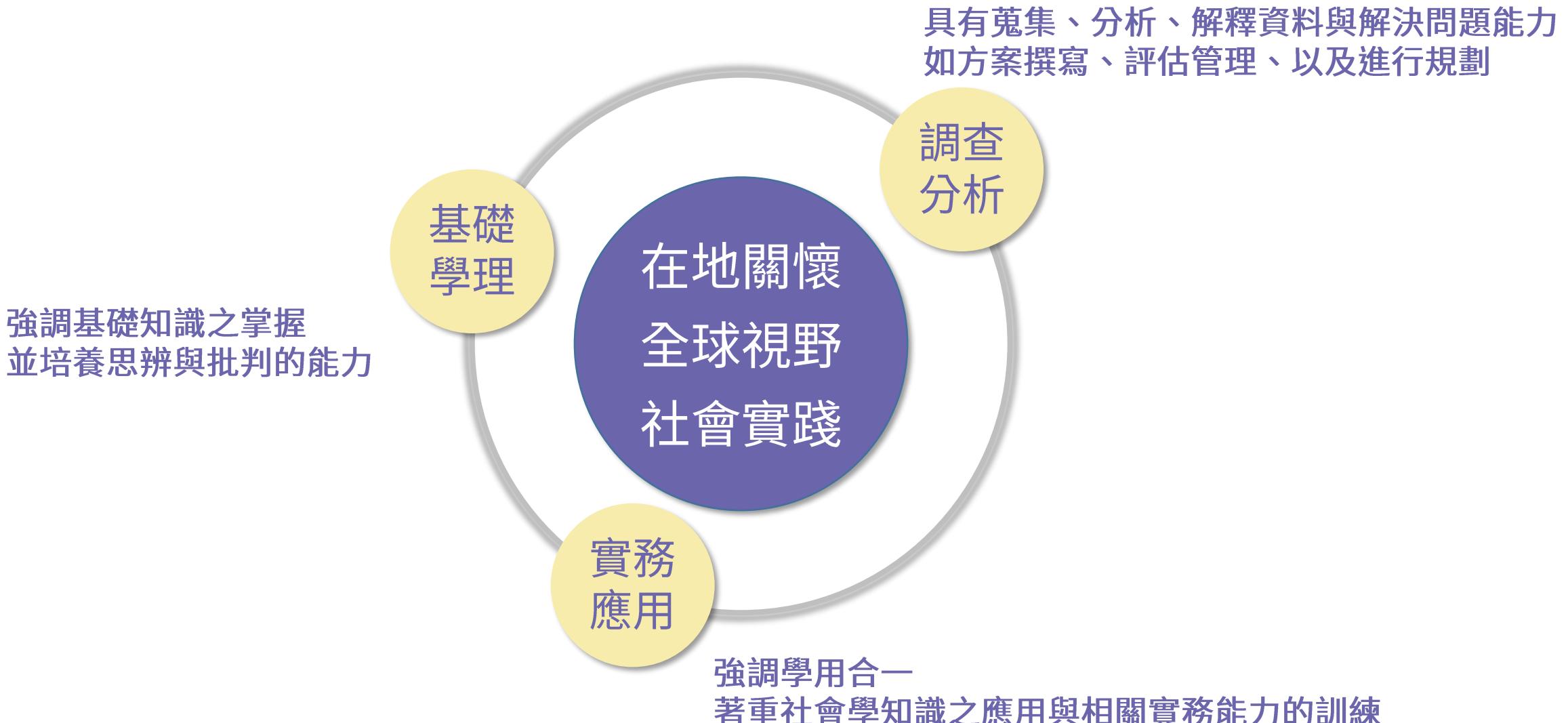
基礎學理、調查分析與實務應用

關於社會學

數據分析

傑華老師的煩惱

總結



Ref: <https://spa.ndhu.edu.tw/p/412-1059-17584.php>

科系、職務與待遇：大數據時代下的社會學與勞動市場

Mail: TLYu0419@gmail.com (#44791)

# 為什麼要教育改革?

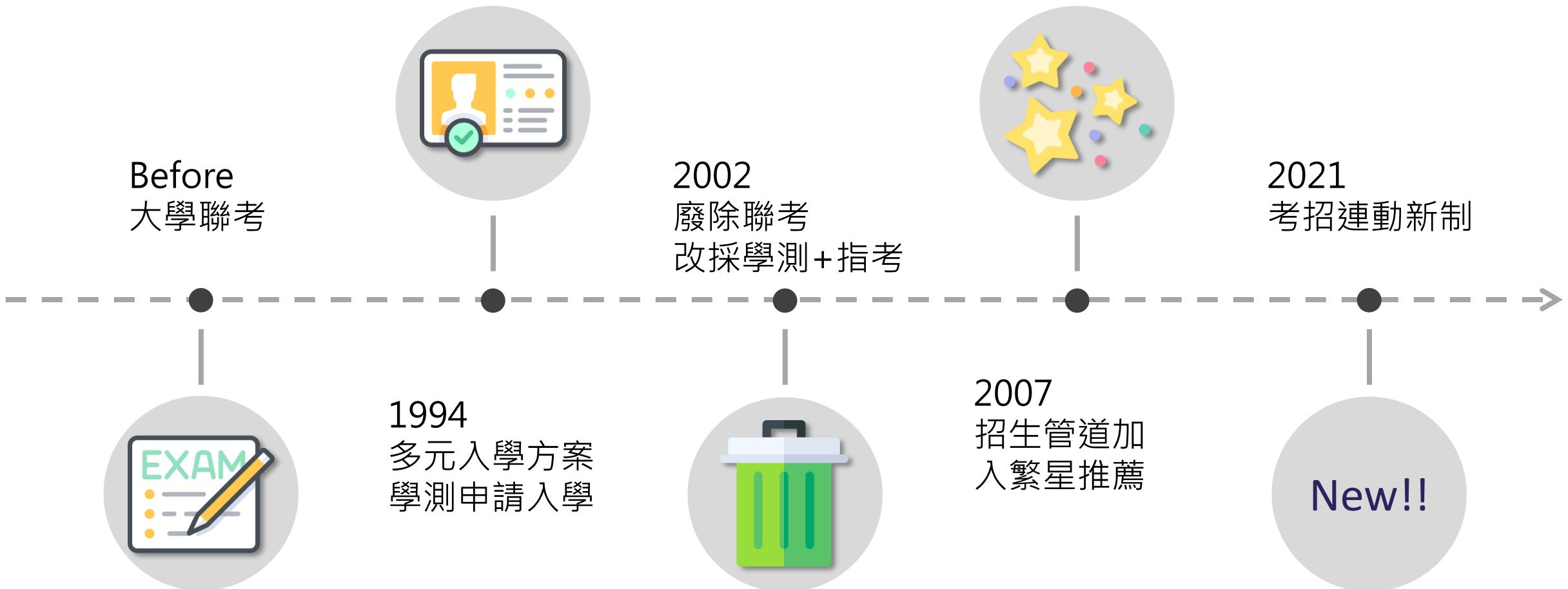
改革是希望教育機會的均等，但也衍生出許多弊端

關於社會學

數據分析

傑華老師的煩惱

總結



- Ref:
1. 駱明慶. (2002). 誰是台大學生?-性別, 省籍與城鄉差異. 經濟論文叢刊, 30(1), 113-147.
  2. 駱明慶. (2018). 誰是台大學生?(2001-2014)-多元入學的影響. 經濟論文叢刊, 46(1), 47-95.

# 為什麼要年金改革？

當財政危機碰上世代正義，背後還牽扯了複雜的政治問題

關於社會學

數據分析

傑華老師的煩惱

總結



拓紹

## 年金改革關鍵數字

	軍公教			勞
退撫基金	軍	公	教	
基金規模	261億元	3390億元	2029億元	勞保6850億元
國庫已撥補 金額	撥補公教保3543.67億元； 軍公教退撫基金34.61億元			勞保及勞工退休金 新制迄今未撥補
未來待撥補 金額	軍保411億元	公教保1313億元		
入不敷出年	2011	2015	2014	2018
破產年	2020	2031	2030	2027
每人平均 月領(註)	4.9萬元	5.6萬元	6.8萬元	勞保1.6萬元(不含 勞退8千~1.3萬元)

註／1、軍公教不含一次領的退休金 2、私校月領1.7萬元、農7256元、國民年金3628元

資料來源／國家年金改革委員會、審計部 製表／陳智華

聯合報

Ref: 1. 年金改革：笨蛋！問題在制度！

2. 臺灣年金制度改革之內涵與思考

3. 軍公教抗議年金改革 明天十萬人上凱道

科系、職務與待遇：大數據時代下的社會學與勞動市場

Mail: TLYu0419@gmail.com (#44791)

# 為什麼會有太陽花學運

## 政黨強行通過兩岸服務貿易協議

關於社會學

數據分析

傑華老師的煩惱

總結



社會學不是反對的社會的經濟成長，  
而是希望討論清楚誰會從中受益，誰又會被犧牲？



Ref: 1. [社會學在衝啥？在照亮黑箱政經結構](#)

2. [〈島嶼天光〉 \(ISLAND'S SUNRISE\) 藝術公民計劃 太陽花運動歌曲](#)

3. <https://zh.wikipedia.org/wiki/太陽花學運>



# 但是畢業之後呢？

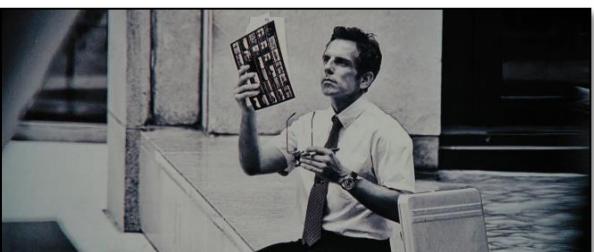
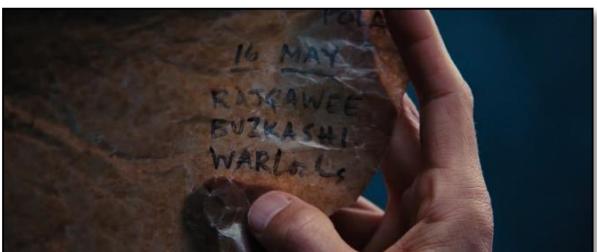
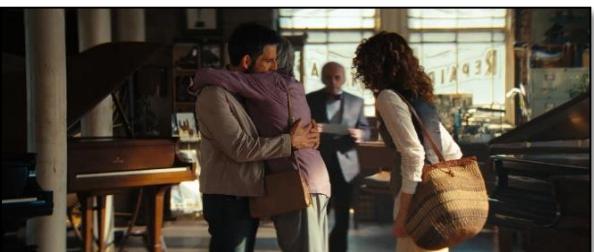
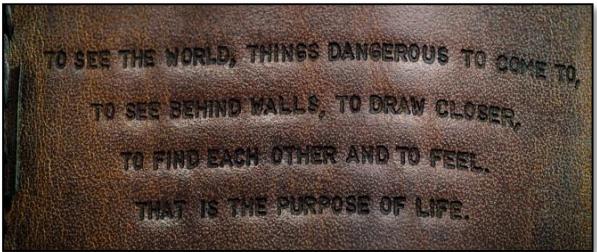


THE SECRET LIFE OF WALTER  
**MITTY**

那唸社會學對職涯有什麼幫助qq

難得的周末我們先來看個影片 輕鬆一下XD







# 那麼職涯規劃呢？

# 關於職涯規劃

看看畢業學長姊們去了哪裡

關於社會學

數據分析

傑華老師的煩惱

總結



National Dong Hwa University Department of Sociology

國立東華大學社會學系

首頁 最新公告 系所簡介 課程規劃 我們的老師 我們的學生 招生資訊 法規及表單下載 新生Q&A 校友風雲榜

首頁 / 招生相關資訊 / 未來大學生參考資料 / 社會學系過來人經驗分享 / 線上同學會

95級-警察-孫英哲 96級-農創-林坤泰 96級-傳播-鍾效京 96級-出版傳播-鍾宛君 96級-社工-張智凱 96級-都計建築-廖一上

96級-設計師-呂孟達 97級-教師-王韋憲 97級-資訊科技-林如珊 98級-非營利組織-顏若卉 98級-教師-黃中一 100級-社工-林思婷

101級-教師-張元鳳 101級-創業-陳貞穎 104級-教師-詹于頃 104級-公職-盧亭榕 104級-社工-鍾政榮 104級-鐵路交通-謝居璋

104級-農漁會-吳博軒 105級-創業-游承輝 105級-公職-賴文婷 105級-人資-林吉偉 105級-公職-曾亞媛 105級-人資-謝佩鈞

105級-斜槓-李英吉 105級-金融保險-王姿雅 106級-警察-古汶玉 107級-人資-蔡忻穎 107級-非營利組織-施芃年 108級-升學-林宜萱

108級-金融-蔡易玲 108級-教師-王慧仙 108級-公職-陳韋誌 108級-升學-蔡佳錚 109級-升學-許冠澤 109級-升學-張利琦

109級-升學-洪紹予 109級-僑生-李文豪 109級大學部經驗分享-黃聖文

Ref: [社會學系過來人經驗分享](#)

科系、職務與待遇：大數據時代下的社會學與勞動市場

Mail: TLYu0419@gmail.com (#44791)

# 關於職涯規劃

看看畢業學長姊們去了哪裡

關於社會學

數據分析

傑華老師的煩惱

總結



National Dong Hwa University Department of Sociology

國立東華大學社會學系

首頁 最新公告 系所簡介 課程規劃 我們的老師 我們的學生 招生資訊 法規及表單下載 新生Q&A 校友風雲榜

首頁 / 招生相關資訊 / 未來大學生參考資料 / 社

95級-警察-孫英哲	96級-農創-林坤泰	96級-
96級-設計師-呂孟達	97級-教師-王韋憲	97級-
101級-教師-張元鳳	101級-創業-陳貞穎	101級-
104級-農漁會-吳博軒	105級-創業-游承燁	105級-
105級-斜槓-李英吉	105級-金融保險-王姿雅	105級-
108級-金融-蔡易玲	108級-教師-王慧仙	108級-
109級-升學-洪紹予	109級-僑生-李文豪	109級-

林如珊 社會發展系97級 國立東華大學社會學研究所碩士  
現職：垣創科技/喬昱科技 專案經理

社會學是一門很廣泛的學科，就因為它的廣泛，所以一般人對於它的印象會覺得廣而不精，但是，就因為此學門含括的範圍極廣，法政經社均有所涉獵，讓我在修讀課程的期間，才更能去摸索探討出自己有興趣的方向去深究研讀它。

研究所期間，有感於自己家鄉-宜蘭-因為雪山隧道的開通，對於交通有了巨大的改變，交通的便利性帶來生活圈的變化，人潮車潮的湧入讓宜蘭許多地方有了明顯的成長，故在研究所就讀時就訂定了論文主題目標。也讓我朝著「都市社會學」這學門開始精進，運用「都市社會學」的核心理論加上「GIS地理資訊系統」為工具，搭配政府的工商業普查統計數據，讓我作出了宜蘭地區在雪隧開通前後五年的都市產業分布的變化。

在論文發表過關之後與辦理離校手續的這段期間，就開始了我的求職面試之旅，每週從花蓮北上台北面試，最後順利面試上台灣第一大GIS原廠資訊公司，也慶幸在研究所期間的學習，讓我帶論文順利錄取與所學相關的工作，更開啟了我從文組背景的學歷意外進入到資訊科技產業領域，也一直發展至今，雖然後來工作的轉職未繼續在GIS產業，但是也讓我意外的深耕了資訊科技產業，至今還是非常感謝系所的多元課綱，讓我更清楚的找到了自己有興趣的發展，並勇於走出自己的路！

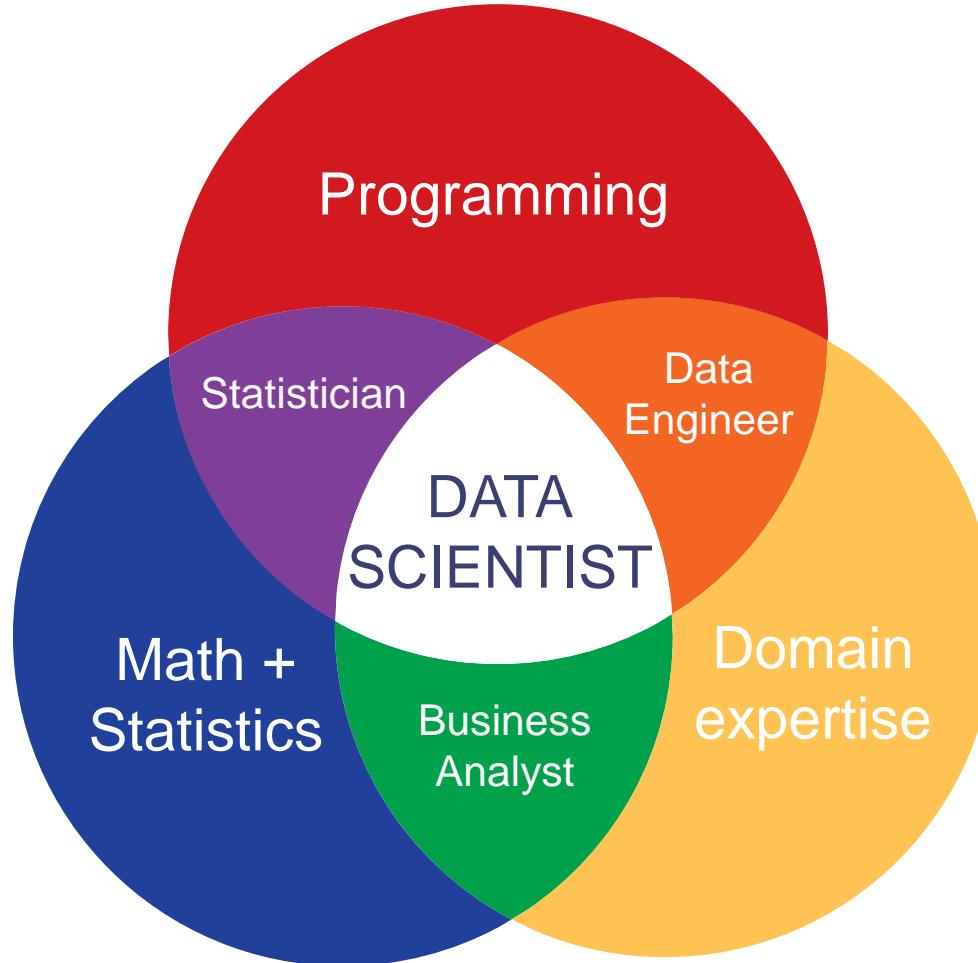
Ref: [社會學系過來人經驗分享](#)

科系、職務與待遇：大數據時代下的社會學與勞動市場

Mail: TLYu0419@gmail.com (#44791)

# 關於職涯規劃

## 資料科學領域技能與職務需求



關於社會學

數據分析

傑華老師的煩惱

總結

### Math + Statistics

- Statistical and Mathematical knowledge (Linear Algebra, Calculus, Metrics and Probability).
- Learn some popular tools like R, Python, and/or Tableau.

### Domain expertise

- Good enterprise expertise helps you build better ML model tools and processes and infer the best business decisions.

### Programming

- Database: Hadoop, SQL and NoSQL
- Cloud Server Architecture: AWS
- toolkits: sklearn, Numpy, Pandas, Keras, etc.
- Algorithms: KNN, LR, NB, RF, GBM, etc.
- ...

Ref: 1. [Data Scientist、Data Analyst、Data Engineer 的区别是什么？](#)

2. [真 · 資料團隊與分工](#)

科系、職務與待遇：大數據時代下的社會學與勞動市場

Mail: TLYu0419@gmail.com (#44791)

# 關於職涯規劃

認識自己才能找到自己的興趣和天賦

關於社會學

數據分析

傑華老師的煩惱

總結



P.S. 這是最後一張劇照了><

- [DISC人格測驗](#)
- [在每個位子上發光：巧用你的工作天賦](#)
- 回想平常在做什麼事情的時候會忙到忘記時間？(遊戲除外)
- 參加業界的參訪/實習
- 來上學長今天的課程

# 關於職涯規劃

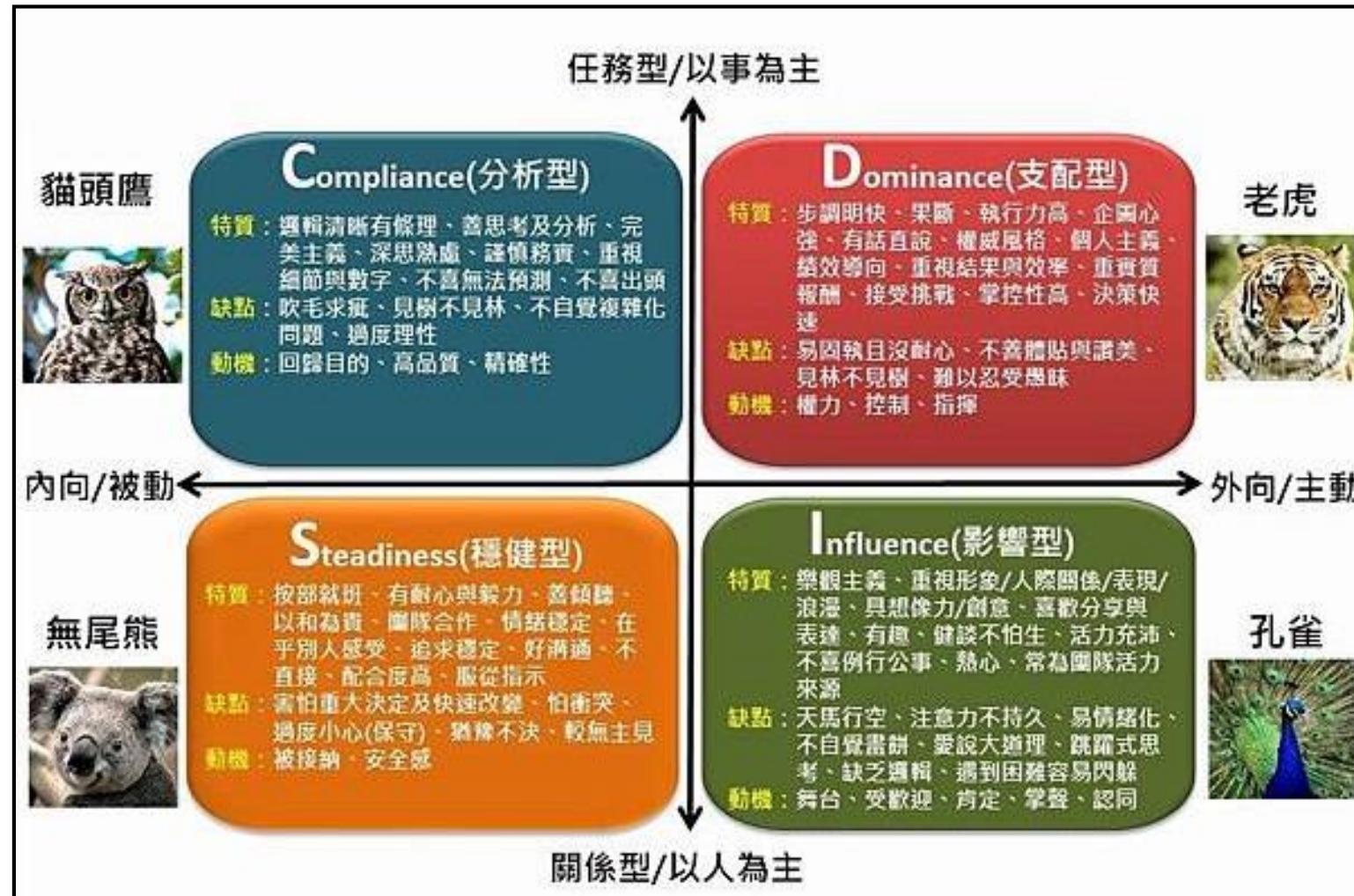
## DISC人格特質測驗說明

關於社會學

數據分析

傑華老師的煩惱

總結



# 休息一下~

- 有問題可以在 Slido 上發問 & 幫別人點讚 (#44791)
- 15分鐘後再回來>"<

先看一下 QA 再繼續 ~

A man in a maroon hoodie and light-colored pants is longboarding down a winding asphalt road. He is leaning into the turn, looking towards the camera. The background features rolling green hills and mountains under a bright, slightly cloudy sky.

# 數據分析

- 分析流程
- 反面應用案例
- 大數據與社會學

# 不就是分析嗎，怎麼還有流程？

回憶一下量化課程做過的事>”<

# 分析流程介紹

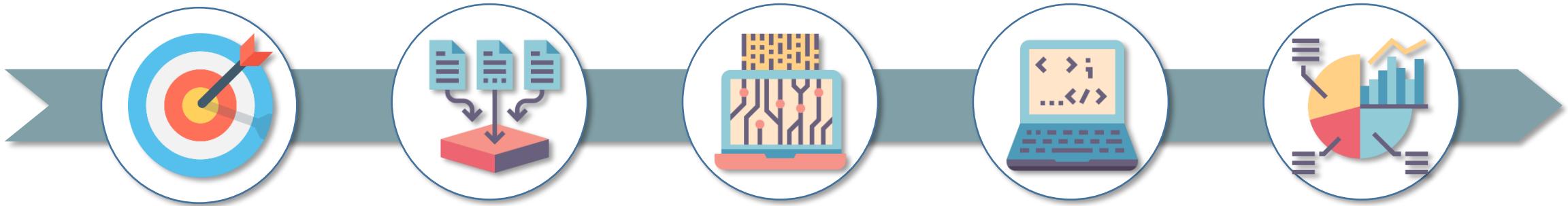
面試時會問，一定要記得很清楚

關於社會學

數據分析

傑華老師的煩惱

總結



## 確認目標

- 重要性
- 定義問題

## 資料蒐集

- 調查問卷
- 公司資料庫
- 公開資料
- 網路爬蟲

## 資料處理

- 資料品質
- 資料清理
- 特徵轉換
- 特徵縮減

## 模型建置

- 分類模型
- 回歸模型
- 分群模型
- 降維模型

## 專案佈署

- 分析專案
- 模型專案

# 確認目標

把問題說清楚，講明白！

## 為什麼這個問題重要？

- 好玩  
預測生存遊戲 [PUBG](#) 誰可以活得久
- 企業核心問題  
如何對用戶更精準的投放廣告 [ADPC](#)
- 公眾利益 / 影響政策方向  
[停車方針](#)、[計程車載客優化](#)
- 對世界很有貢獻  
[肺炎偵測](#)

關於社會學

數據分析

傑華老師的煩惱

總結

## 要回答什麼問題？

- 問題要能夠被轉化為具體的指標
- 根據目標變數型態會需要使用不同的衡量指標
- [PUBG](#)：生存時間、遊戲排名
- [ADPC](#)：廣告點擊率、商品成交率
- [肺炎偵測](#)：影像辨識的準確度
- ...

# 確認目標

把問題說清楚，講明白！

關於社會學

數據分析

傑華老師的煩惱

總結

## 為什麼這個問題重要？

- 好玩

預測生存遊戲 PUBG 誰可以活得久

- 企業核心問題

如何對用戶更精準

- 公眾利益 / 影響

停車方針、計程車

- 對世界很有貢獻

肺炎偵測

## 要回答什麼問題？

- 問題要能夠被轉化為具體的指標
- 根據目標變數型態會需要使用不同的衡量地標

類別	數學特徵	範例
類別	=, ≠	居住縣市，科系名稱...
順序	=, ≠, >, <	滿意度，消費頻率...
等距	=, ≠, >, <, +, -	溫度，年份...
等比	=, ≠, >, <, +, -, *, /	營業額，來客數...

# 確認目標

把問題說清楚，講明白！

關於社會學

數據分析

傑華老師的煩惱

總結

知乎 | 首发于 王喆的机器学习笔记

2019-05-05

宁肯  
一个清华大学的，人生价值在于研究让人更多地点击广告，这不是很可悲的一件事吗？  
1

12 条回复

王喆 (作者) 回复 宁肯  
清华大学一年只要出十个为国为民的大家就够了，其他人做一些养家糊口的事情挺好  
49

行者 回复 宁肯  
对于一个普通人来说，好好工作就是对国家和社会的最大贡献；而且和答主相比，我觉得大多数人都是造粪机器（比如我）

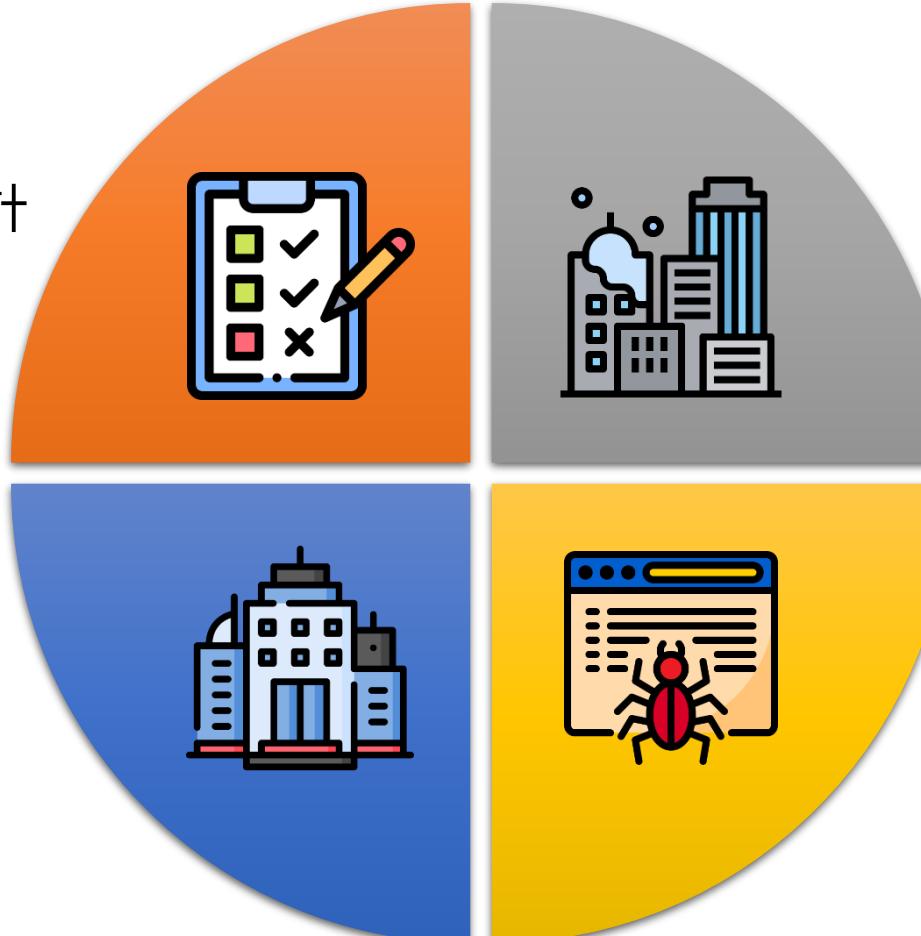
5

Ref: 深度学习中不得不学的Graph Embedding方法

### 調查問卷

依專案需求，由領域專家設計與發放問卷的方式蒐集資料

- [台灣社會變遷基本調查](#)



### 政府公開資料

政府部門提供之公開資料

- [經濟地理資訊系統](#)
- [捷運各站進出量統計](#)

### 公司資料庫

依公司個別業務累積的客戶行為、消費、位置……等等資料

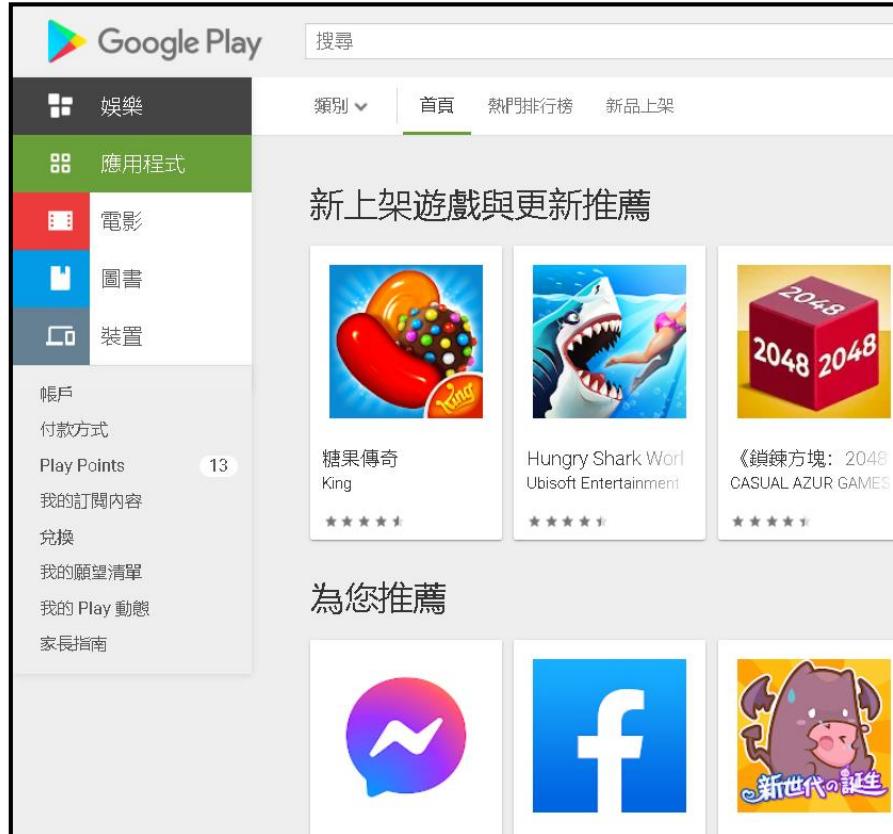
### 網路爬蟲

透過網路爬蟲程式擷取資料

- [GooglePlay](#)
- [松果購物](#)

# 資料蒐集

資料在哪裡？要怎麼取得資料？



APP 名稱

Facebook

Facebook 社交

★★★★★ 104,679,238

含廣告內容 · 提供應用程式內購  
這個應用程式與你所有的裝置都相容。

加入願望清單

安裝

整體滿意度

評論

4.2

★★★★★ 104,679,238 則評分

客戶名稱、時間、個人滿意度、評論內容、被讚數

jolin shen ★★★★★ 2020年11月4日

從9月中以後，廣告幾乎是沒有效果，我問其他同樣管理粉絲專頁的人，大家都是同樣反應 請問為何繳了費用，為何廣告效益差那麼多？台灣沒有疫情，沒有封城，大家都正常上班上課 是FB系統在轉換或更新嗎？如果是如此，是否可以歸還廣告費用呢？

30

Ref: <https://play.google.com/store/apps>

科系、職務與待遇：大數據時代下的社會學與勞動市場

Mail: TLYu0419@gmail.com (#44791)

# 資料蒐集

我想知道電商的價格與客戶的評論

關於社會學

數據分析

傑華老師的煩惱

總結

The screenshot shows the Shopee website interface. At the top, there are navigation links for '賣家中心' (Seller Center), '下載' (Download), '追蹤我們' (Follow us), and social media icons. The main search bar has a placeholder '看更多11.11免運商品' (More 11.11 free shipping products). Below the search bar, there are two promotional banners: one for the '11.11 最強購物節' (11.11 Strongest Shopping Festival) with offers like '全面11折' (All 11% off) and '天天0免運 \$11起標豪宅' (Daily 0 free shipping, starting from \$11); and another for '筆記型電腦' (Laptops) with a large blue header box containing the text '商品名稱、價格、滿意度、銷售量' (Product name, price, satisfaction, sales volume).

**條件篩選 (Conditions Filter):**

- 分類 (Category):
  - 電腦周邊配件 (5萬+)
  - 電腦零組件 (2萬+)
  - 鍵盤滑鼠 (2萬+)
  - 居家生活 (2萬+)
- 更多 (More)
- 出貨地點 (Shipment Location):
  - 台灣
  - 海外
- 運送方式 (Delivery Method):
  - 全家
  - 7-11
  - 黑貓宅急便
  - 萊爾富
- 更多 (More)
- 品牌 (Brand):

**筆記型電腦 (Laptops) Search Results:**

商品名稱	價格範圍	滿意度	銷售量	出貨地點
15.6吋 筆電 win10 超速 戰勝 asus acer imac 打遊...	\$3,499 - \$8,999	★★★★★	已售出 147	新北市永和區
MITPC*acer宏碁 S3 391 UltraBook 超輕薄筆電 ...	\$5,500	★★★★★	已售出 473	台中市北屯區
14吋 全新筆電 win10 超速 快感 戰勝 asus 筆電 ipa...	\$2,999 - \$8,999	★★★★★	已售出 381	新北市永和區
15.6吋 win10 打lol 英雄聯盟 戰勝 asus 筆電 筆記...	\$3,499 - \$7,999	★★★★★	已售出 127	新北市永和區
ASUS 15.6吋 i5 4G/240G SSD K55A Win10二手...	\$4,600 - \$5,100	★★★★★	已售出 127	新北市永和區
ASUS X205TA 珍珠白 WIN10	\$4,500	★★★★★	已售出 19	新北市中和區
ASUS 華碩 E410MA [... 滿額贈]	\$10,900 - \$9,999	★★★★★	已售出 2	台北市中正區
清倉【ASUS】華碩 S410UA (i5-...	\$13,500	★★★★★	已售出 1	彰化縣彰化市
14吋 全新筆電 win10 超速 快感 全新 中文 筆電 筆...	\$2,999 - \$7,999	★★★★★	已售出 186	新北市永和區
3C客製筆電 i7 i5 intel AMD獨顯 筆電 筆記型...	\$4,500 - \$13,000	★★★★★	已售出 141	桃園市八德區

Ref: <https://shopee.tw/>

科系、職務與待遇：大數據時代下的社會學與勞動市場

Mail: TLYu0419@gmail.com (#44791)

# 資料蒐集

## 我想知道候選人的動態與網路聲量

The screenshot shows a Facebook profile for Tsai Ing-wen. On the left, there's a post from 9 hours ago about attending the Golden Horse Film Festival (TGHFF) and the impact of the pandemic. On the right, there's a larger image of Tsai speaking at a podium with microphones from various media outlets. Below the image is a highlighted post from 1 hour ago where she discusses民調 (public opinion poll) and its relation to the will of the people. The post has received 17K likes, 916 comments, and 267 shares. A yellow box highlights the comment section below this post, which contains three comments from Tsai Ing-wen and one from Lü Yu-chang.

關於社會學 數據分析 傑華老師的煩惱 總結

蔡英文 Tsai Ing-wen 9 小時  
因為疫情的影響，我們的生活有了一些改變，但常進行，一樣能走進電影院、舉辦頒獎典禮。前陣子，我參與了金馬影展 TGHFF 的社群活動。金馬獎，是每年台灣的電影盛事，相信喜愛電影在還在院線，而且有入圍今年金馬獎的台灣電影... 查看更多

林雨蒼和其他 6.1 萬人

不論是民心還是民調，民主國家的人民，有自己自由的意志。民調好，更要全力衝刺；民調不好，就深刻檢討。

通關密語 1203 記得收看我的直播

關於社會學 數據分析 傑華老師的煩惱 總結

蔡英文 Tsai Ing-wen 1 hr  
不論是民心還是民調，民主國家的人民，有自己自由的意志，不是政治人物可以下令要人民做什麼事情，政治人物只能拜託人民，為人民做事情。

#我的通關密語1203  
#鎖定直播

Like Comment Share

Most Relevant

Author 蔡英文 Tsai Ing-wen https://facebook.com/events/473640546595775/?ti=icl Like Reply 2h 141

Author 蔡英文 Tsai Ing-wen https://www.facebook.com/events/473640546595775/?ti=icl Like Reply 2h 101

呂玉產 之前一位中國人在問，中國已經是全球的第二經濟體了，為什麼台灣人不願意接受一國兩治，我的回答是，中華民國已經是主權國家，沒有二元政權存在。

Write a comment...

貼文人、  
貼文時間、  
貼文內容、  
按讚數、  
留言、分享數

留言人、  
留言時間、  
留言內容、  
按讚數、  
回覆內容

Ref: <https://www.facebook.com/tsaiingwen>

科系、職務與待遇：大數據時代下的社會學與勞動市場

Mail: TLYu0419@gmail.com (#44791)

# 資料預處理

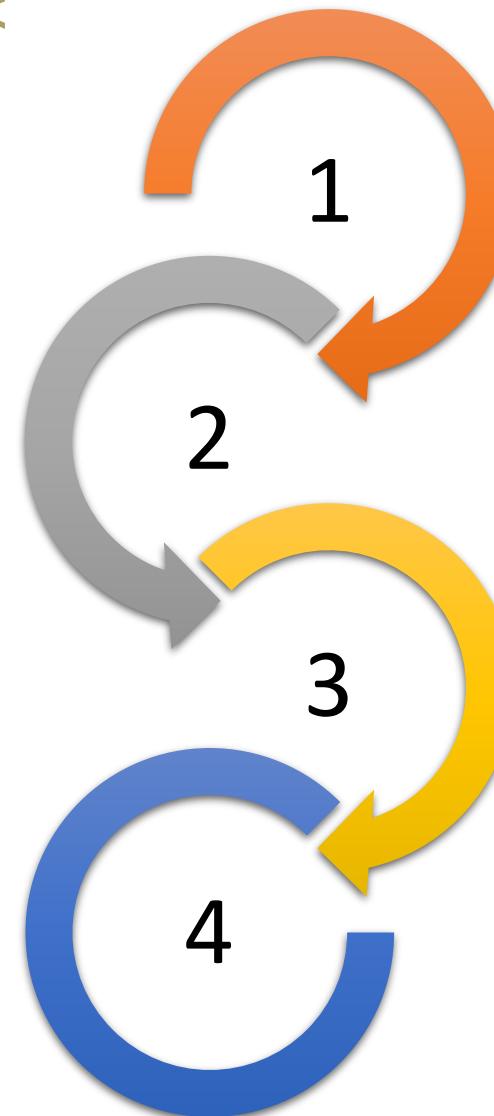
其實這些知識你們都學過>“<

## Data reduction

1. Attribute feature selection
2. Dimensionality reduction
3. Numerosity reduction

## Data cleaning

1. Missing data
2. Noisy data



## Data quality assessment

1. Mismatching in data types
2. Different dimensions of data arrays
3. Mixture of data values
4. Outliers in the dataset
5. Missing data

## Data transformation

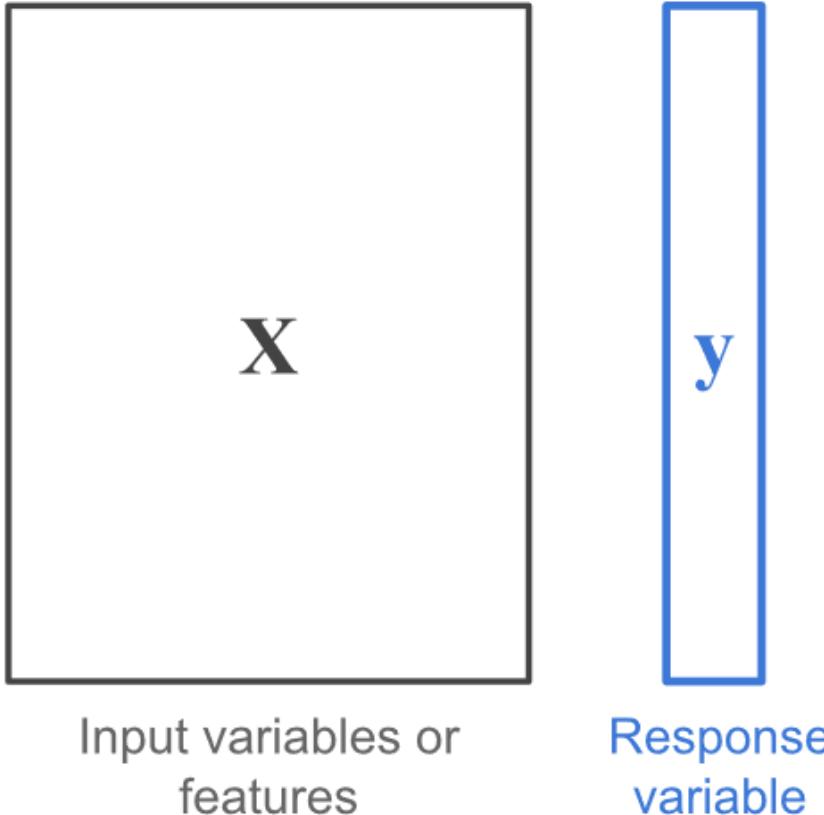
1. Aggregation
2. Normalization
3. Feature selection
4. Discretization
5. Generalization

Ref: <https://serokell.io/blog/data-preprocessing>

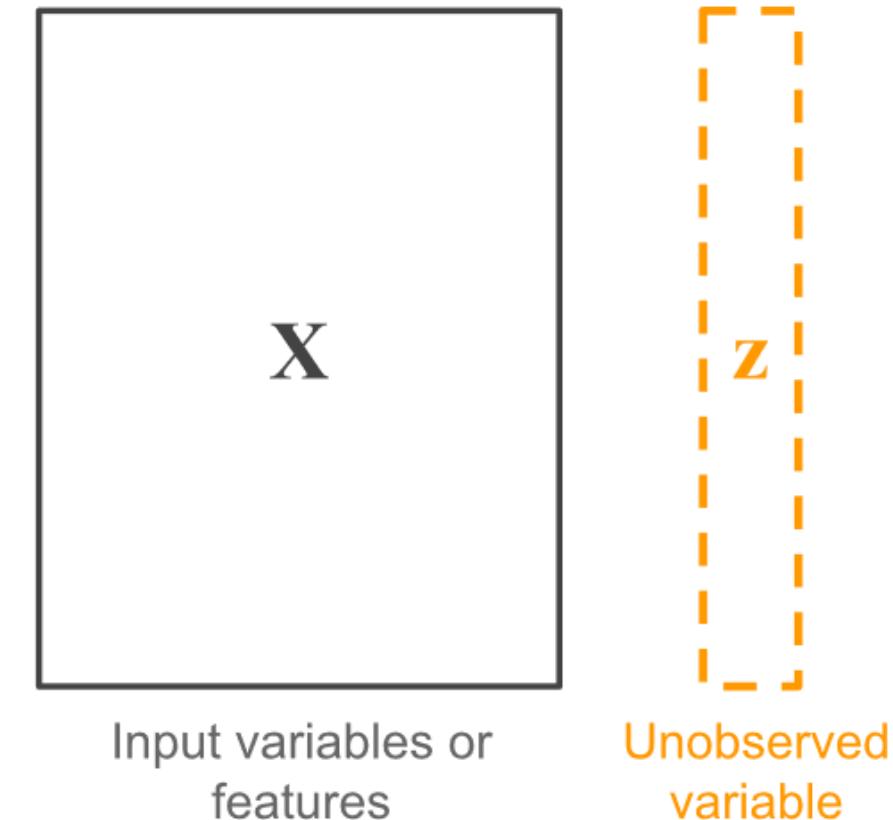
科系、職務與待遇：大數據時代下的社會學與勞動市場

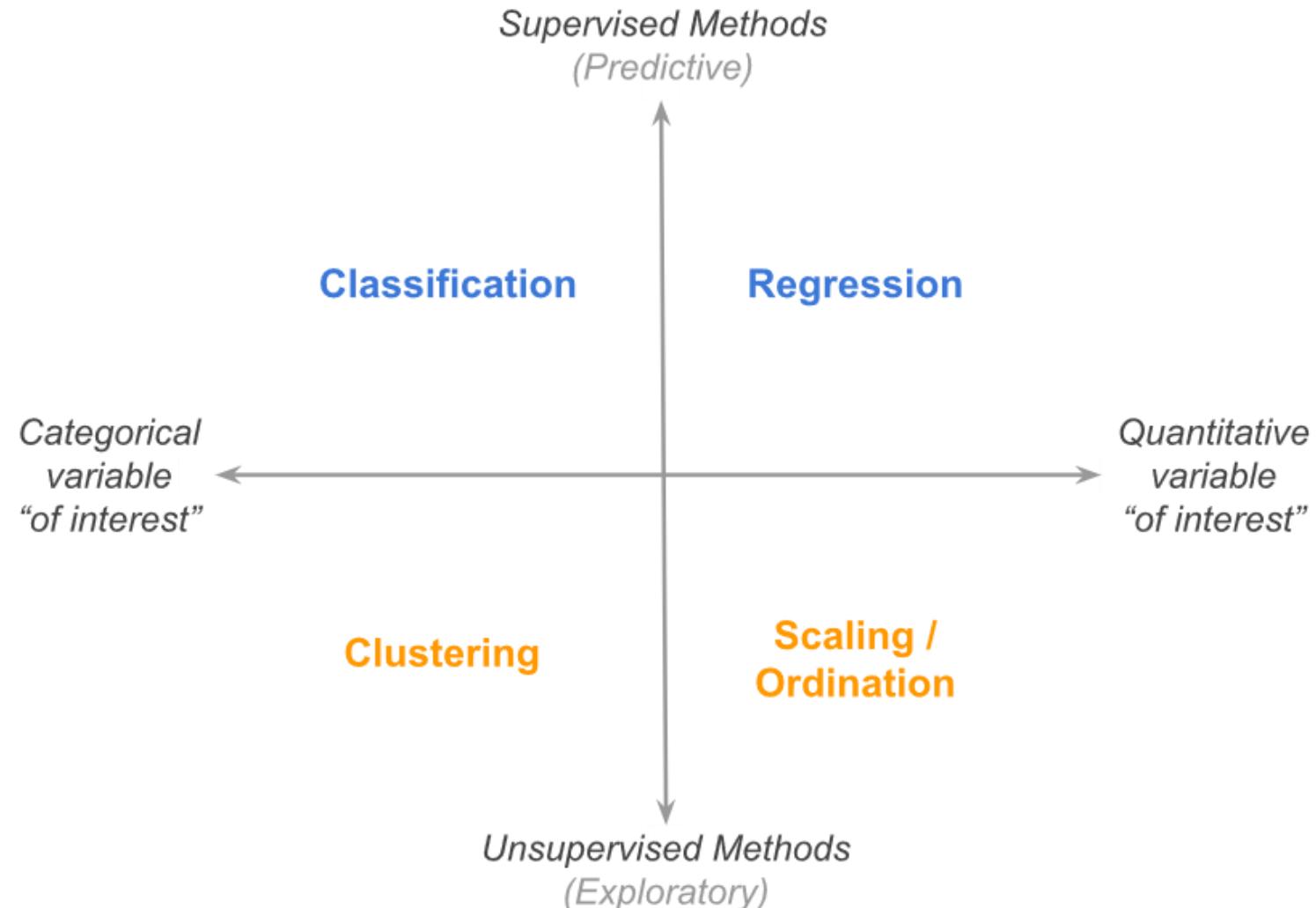
Mail: TLYu0419@gmail.com (#44791)

### *Supervised Learning*



### *Unsupervised Learning*





# 模型建置

## 分類模型應用案例

關於社會學

數據分析

傑華老師的煩惱

總結

- 預測學生延畢
- 預測是否會升學
- 預測是否會找到工作
- ...



學校怎麼用

分類模型  
Classification

企業怎麼用

- 商品購買推薦
- 客戶流失風險
- 客戶進線問題預測
- ...



當預測的目標是類別變數就屬於分類模型，  
依目標變數的類別數量可以分為二分類與多分類模型

# 模型建置

到底什麼是型一型二錯誤>"<

關於社會學

數據分析

傑華老師的煩惱

總結

Type I Error



Type II Error



只要是預測就會有犯錯的可能，  
不同的犯錯情況會衍生出犯錯的成本

- 學年成績預測
- 預測多久會畢業
- 預測多久能找到工作
- ...



學校怎麼用

回歸模型  
Regression

企業怎麼用



- 消費金額預測
- 通話時長預測
- ...

當預測的目標是數值變數就屬於回歸模型，  
相較於分類模型可以更細緻的捕捉到程度上的差異

# 模型建置

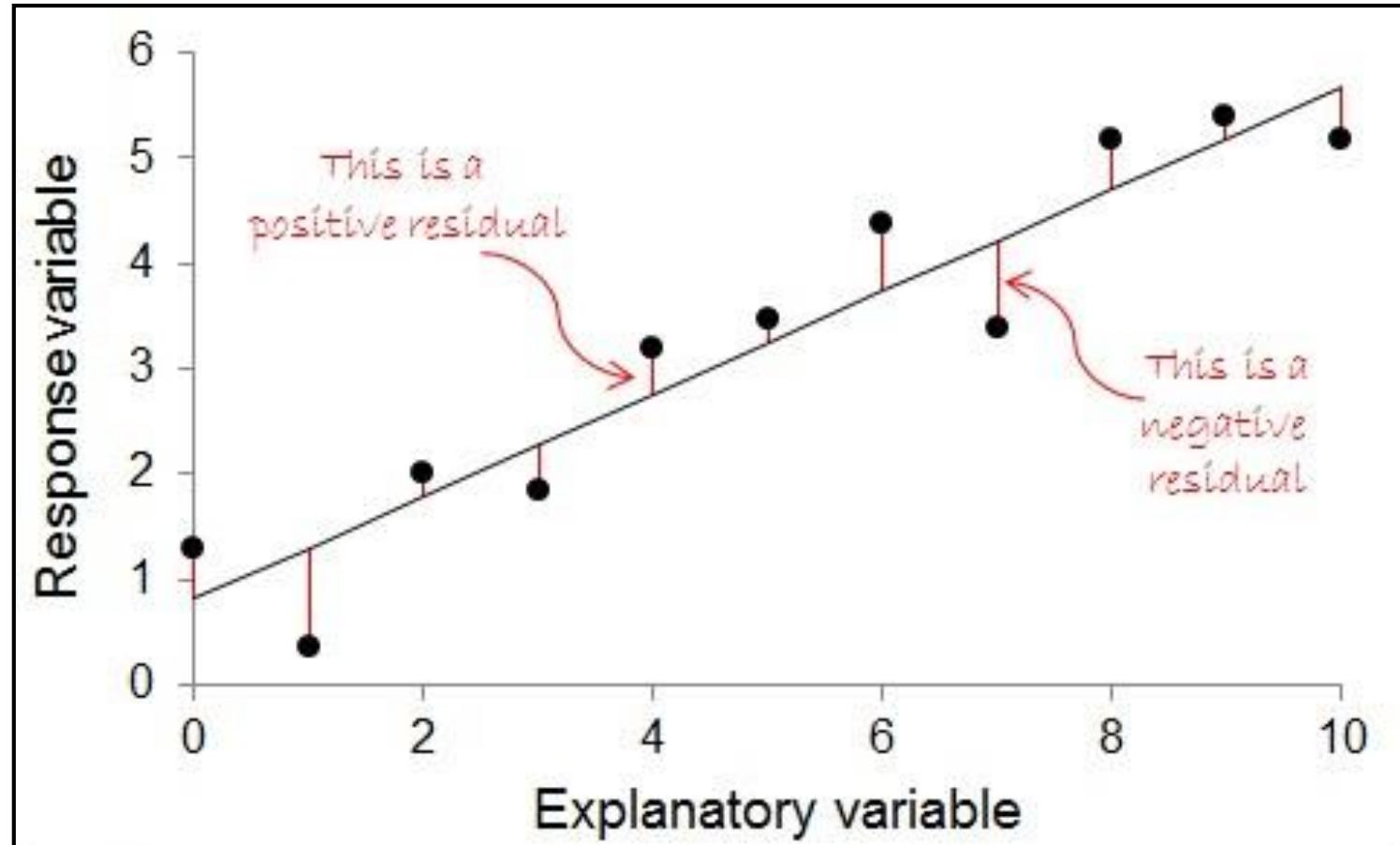
高估與低估都有各自因犯錯帶來的成本

關於社會學

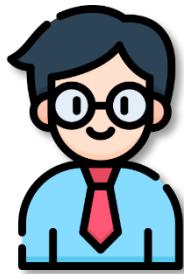
數據分析

傑華老師的煩惱

總結



- 音樂品味分析
- 運動偏好分析
- 閱讀習慣分析
- .....



學校怎麼用

降維模型  
Dimension Reduction

企業怎麼用

- 消費習慣分析
- 網頁瀏覽行為分析
- 客戶進線問題分析



以基本學力測驗的科目為例，  
我們會有國、英、數、地理、歷史、公民、物理、化學…等等成績，  
透過降維模型可以幫我們找出地理、歷史、公民是高度相關的科目(社會科)，  
物理、化學是高度相關的科目(自然科)，我們能藉此萃取出潛在結構

# 模型建置

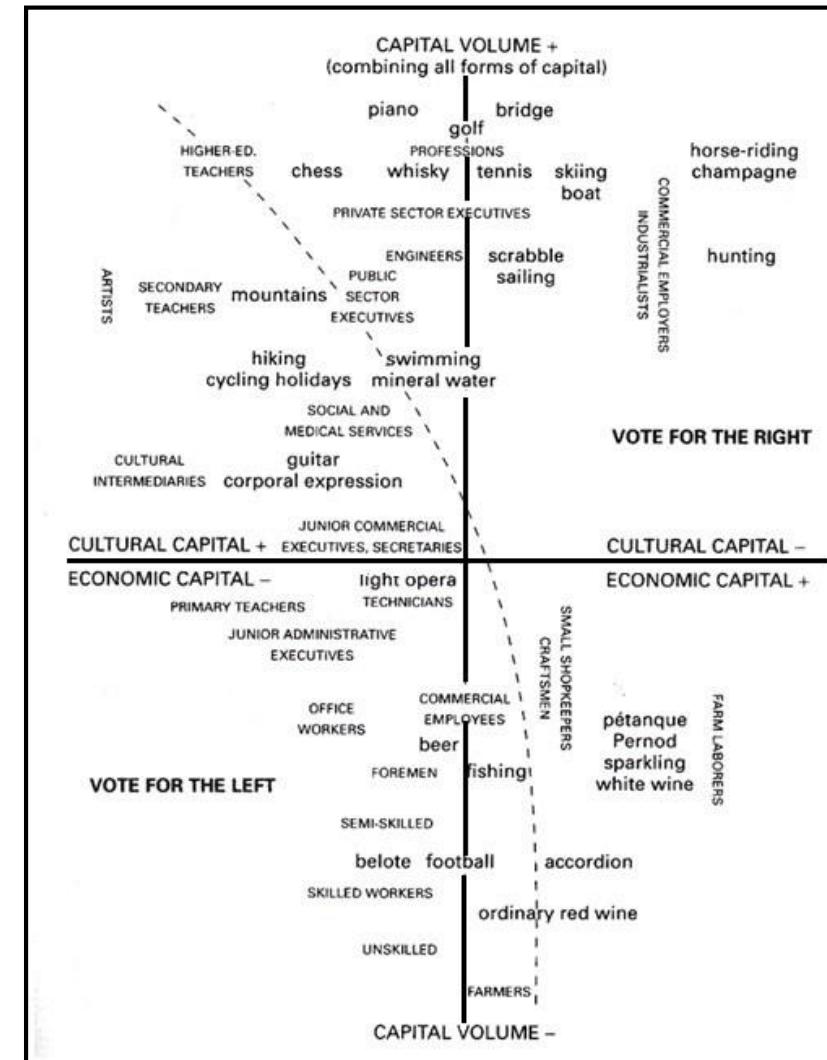
當維度較多時可以透過降維技術提取潛在結構

關於社會學

數據分析

傑華老師的煩惱

總結



- 成績分群
- 生活習慣分群
- .....



學校怎麼用

分群模型  
Clustering

企業怎麼用

- 欠費行為分群
- 消費行為分群
- 進線行為分群



降維模型中找到的潛在結構會反應程度上的差異，  
而分群模型可以直接跟我們說分成幾群，各群樣貌的差異  
(但同時也會丟掉一些「程度」上的資訊)

# 模型建置

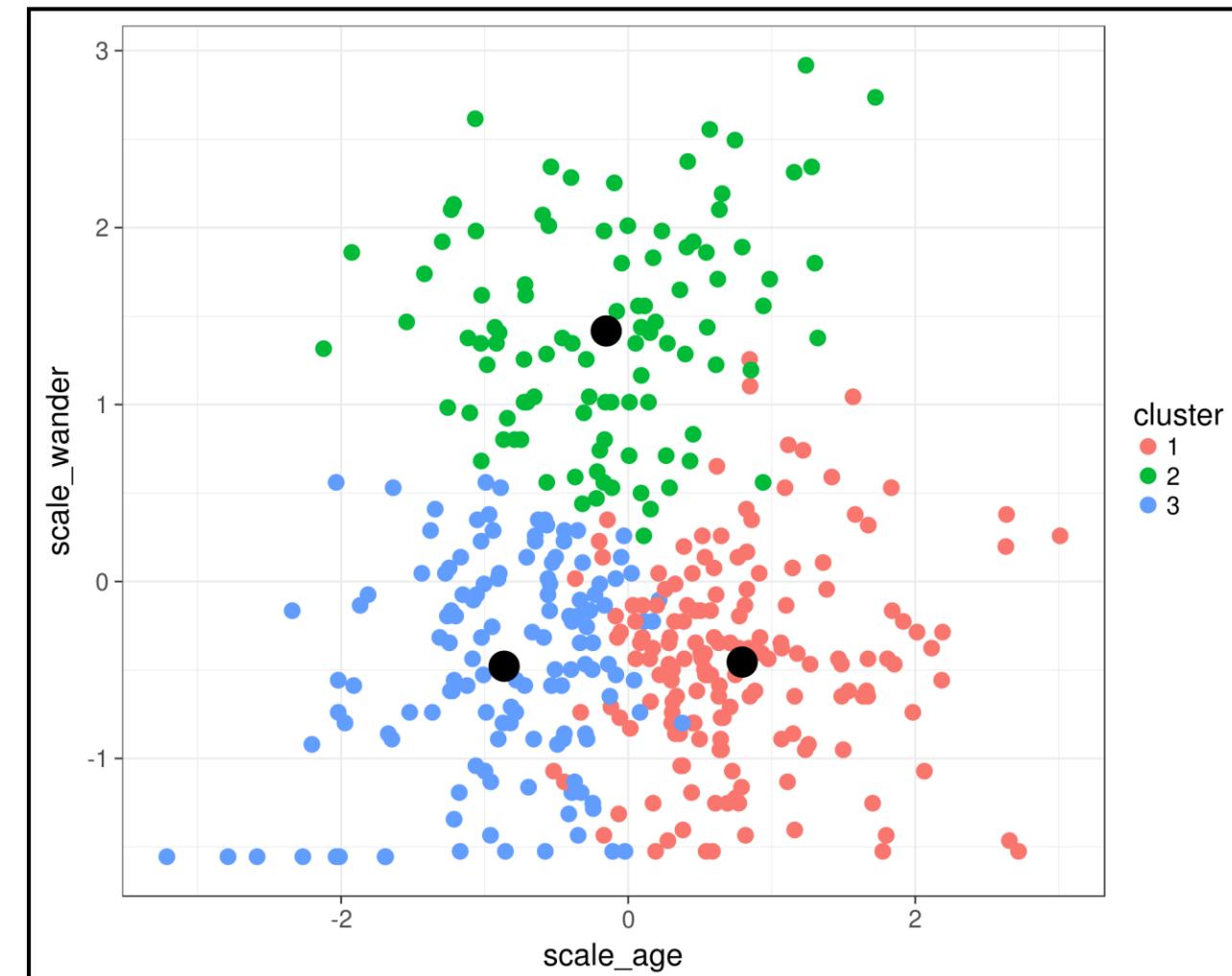
分群技術能幫我們快速理解分群組數與特色

關於社會學

數據分析

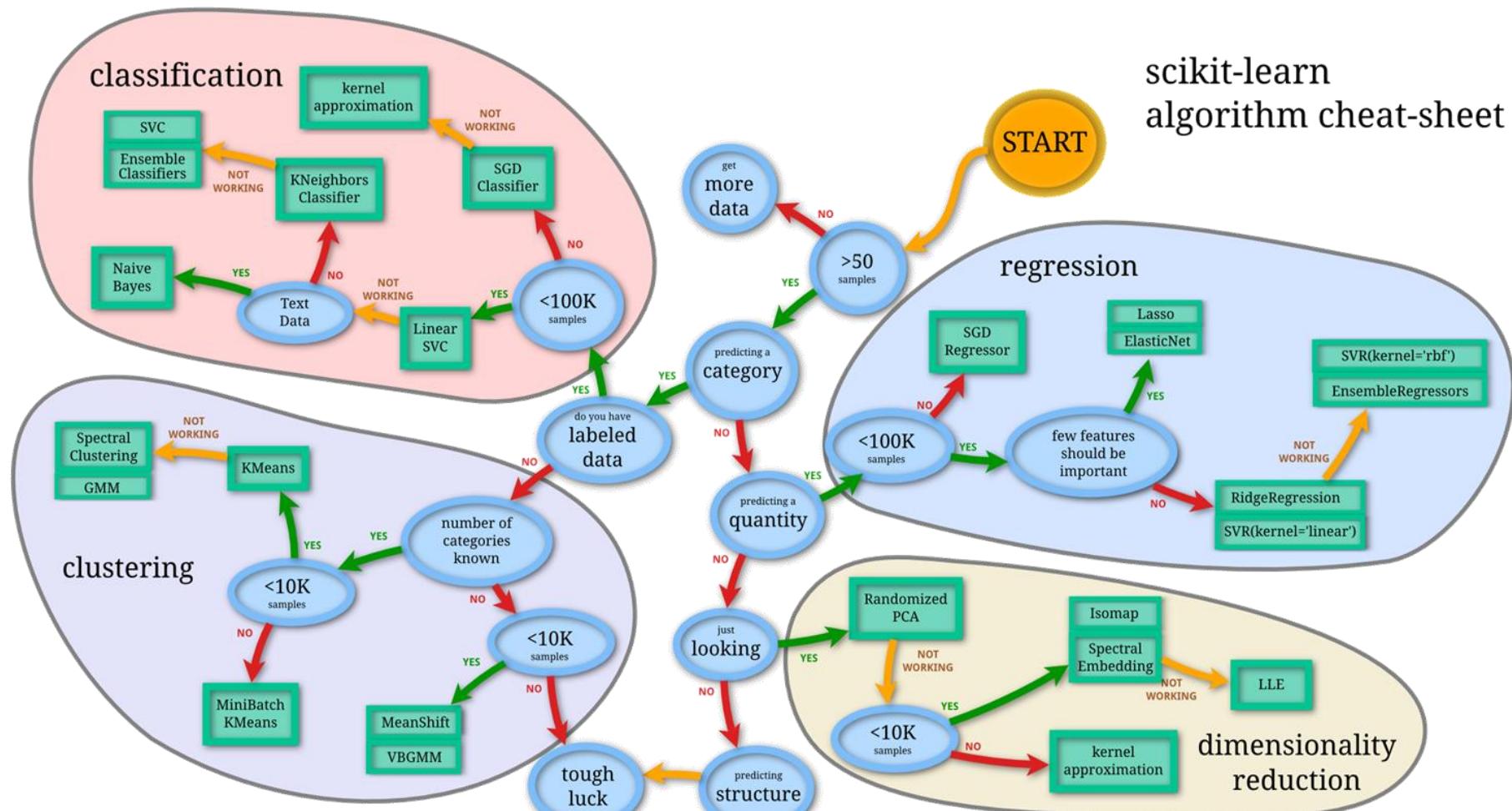
傑華老師的煩惱

總結

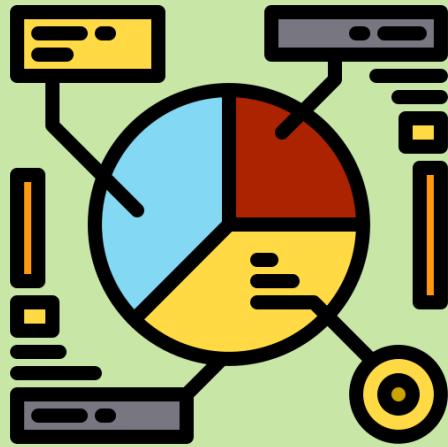


# 模型建置

## 如何選擇合適的演算法



### 分析案



- 透過視覺化的技巧，協助長官理解想傳遞的想法
- 要有清楚、具體的建議，與後續的工作項目

### 模型案

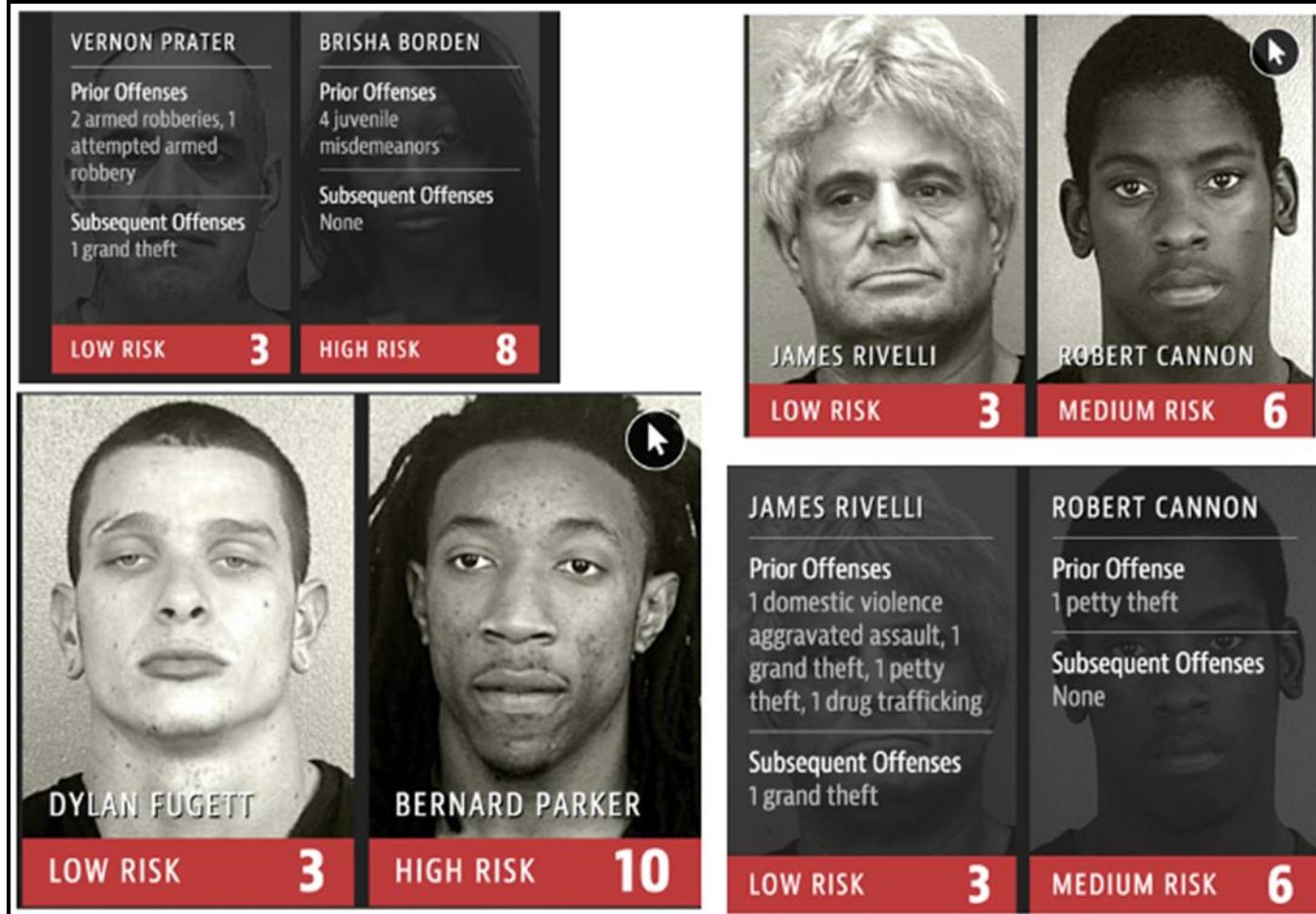


- 建立預測模型  
精準的找到目標客戶或訂定目標
- 要清楚的告訴業務單位改變跟影響是什麼
- 需要設計追蹤的機制定期檢視模型成效

# 反面應用案例

# 反面應用案例

## 隱藏在大數據中立客觀下的歧視



- 歧視被包裹在「技術中立」的外衣中
- 運用大數據預測未來犯罪機率的「風險評估數字」呈現對黑人的嚴重偏見

Ref:1. [Machine Bias](#)

2. [大數據應用潛藏的歧視風險](#)

# 反面應用案例

## 精準推薦反而促成網路成癮

關於社會學

數據分析

傑華老師的煩惱

總結



網路平台成功的預測了用戶的偏好  
但也在因此助長了網路成癮的問題

Ref: 網路成癮大解密！從何而來又該如何解決？

科系、職務與待遇：大數據時代下的社會學與勞動市場

Mail: TLYu0419@gmail.com (#44791)

# 反面應用案例

## 精準推薦反而促成網路成癮

關於社會學

數據分析

傑華老師的煩惱

總結

各平台成功的預測了用戶的偏好  
也在因此助長了網路成癮的問題

Ref: 網路成癥大解密！從何而來又該如何解決？

科系、職務與待遇：大數據時代下的社會學與勞動市場

Mail: TLYu0419@gmail.com (#44791)

53

# 反面應用案例

透過網軍帶風向嘗試影響選民態度

關於社會學

數據分析

傑華老師的煩惱

總結

INSIDE 5G AI 新創 評論 焦點 活動 Jobs 好工作 繁 / 簡 訂閱 :

趨勢

## 抓到網軍！Twitter 永久刪除來自中、俄、土耳其的 3.2 萬個帳號

2020/06/12 · Heemie · Twitter、網軍、言論、仇恨言論

除了刪掉中國核心網軍的大量帳號，Twitter 也發現有額

軍帳號所發表內容的正當性。



▲ Photo Credit: Shutterstock/達志影像

中時電子報 真道理 真愛台灣

新冠肺炎 即時 政治 言論 生活 娛樂 財經 社

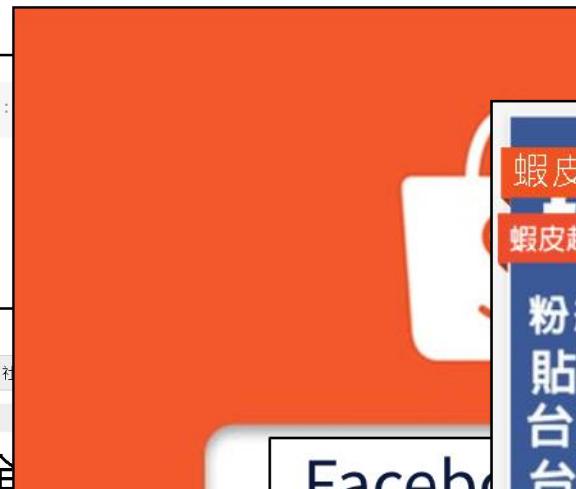
首頁 / 中國時報

## 網軍帶風向價碼 網站全

04:10 2020/06/26 | 中國時報 | 吳家豪



拍賣網站驚見網軍出售帶風向的服務，且價格全都露。圖為國民黨去年率眾到行政院陳情，要求消滅網軍東廠。（本報資料照）



蝦皮優選  
蝦皮超便宜

facebook

粉絲專頁100(讚+追蹤)  
貼文100讚 40元  
台灣真人50讚 50元  
台灣真人留言讚 2.8元  
直播 LIVE 80元

全館最便宜

全店免運

facebook 按讚 / 直播 / 真人 / 留言讚 / 粉絲專頁讚 / 直...

\$10 - \$50

Free

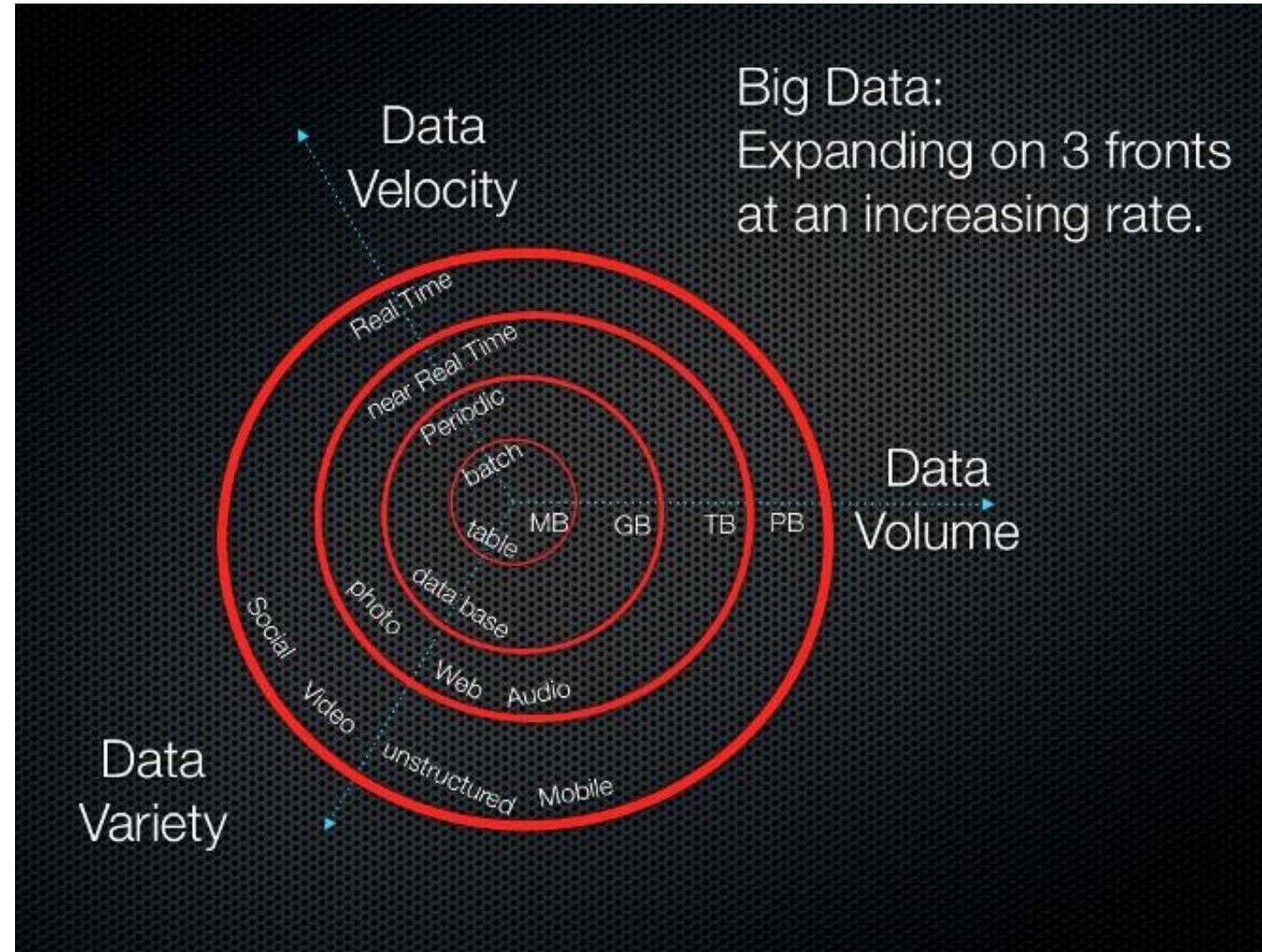
月銷量 1 萬

台南市佳里區

# 到底大數據是好還是壞>“<

# 大數據與社會學有什麼關係？

## 大數據的3個V: 資料量、傳輸速度、資料多樣性



# 大數據給社會學的挑戰

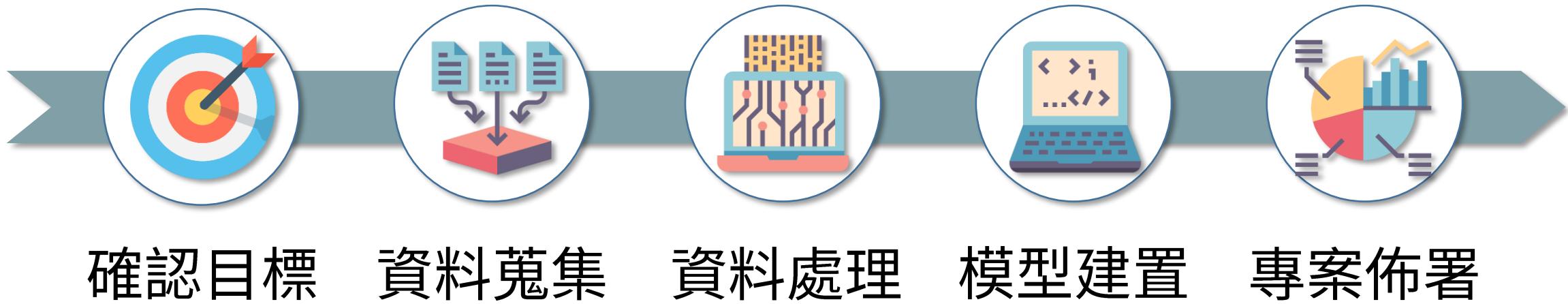
## 回來看一下分析流程

關於社會學

數據分析

傑華老師的煩惱

總結



# 大數據給社會學的挑戰

回來看一下分析流程

關於社會學

數據分析

傑華老師的煩惱

總結

	社會學	大數據
確認目標	從社會理論出發， 找出社會的不平等並設法改善	從數據角度出發， 設法改善業務流程、提升銷售量
資料蒐集	以問題意識為核心來設計與蒐集問卷，知識密度較集中，資料品質高	蒐集資料時通常沒有特定的目的， 成本較低，但資料碎片化問題嚴重
資料處理	結構化的數據	結構化的數據+非結構化數據
模型建置	線性回歸、邏輯回歸	機器學習、深度學習
專案部屬	用嚴謹的研究設計來驗證/修正知識 重視正確性與因果關係	講究快速解決問題

Ref: [大数据给社会学研究带来了什么挑战？邱泽奇](#)

科系、職務與待遇：大數據時代下的社會學與勞動市場

Mail: TLYu0419@gmail.com (#44791)

# 大數據給社會學的挑戰

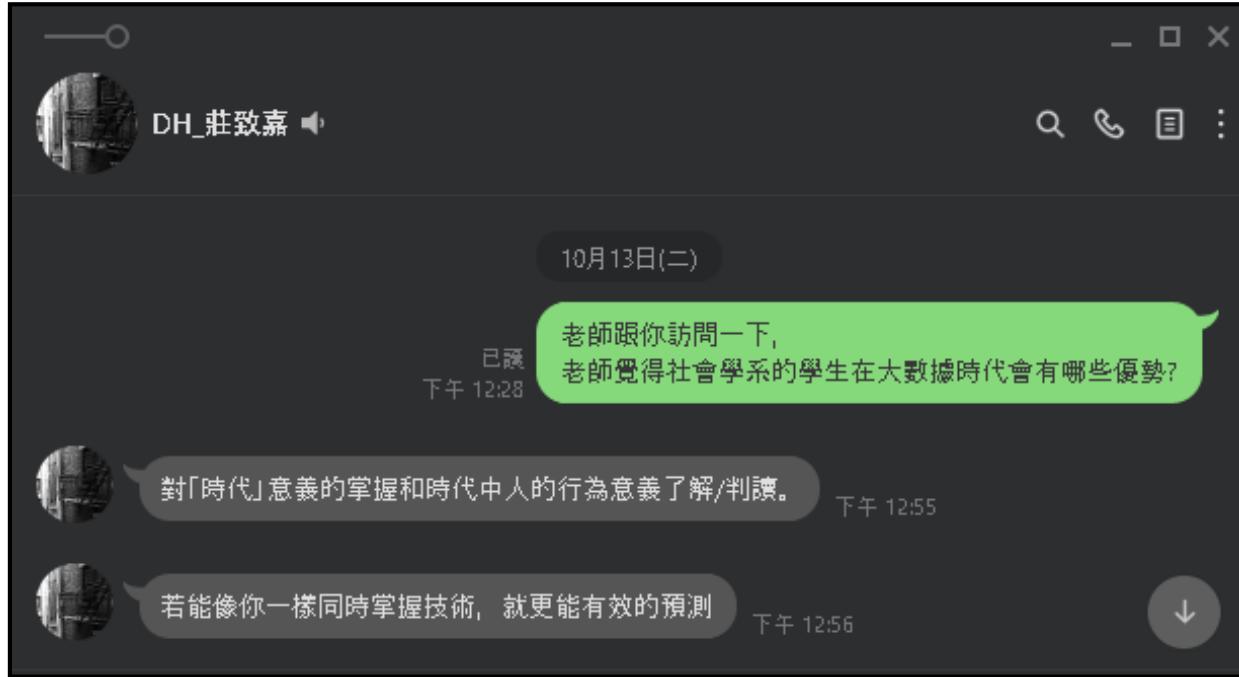
有請致嘉老師協助回答xD

關於社會學

數據分析

傑華老師的煩惱

總結



↑

2020-10-13的訪談>"<

## • 社會學的核心

- 調查資料
- 分析工具
- 社會理論的知識

## • 建議

- 加強資料蒐集的能力
- 提升分析技能
- 留意社會變遷趨勢（資訊化社會）

# 休息一下~

- 有問題可以在 Slido 上發問 & 幫別人點讚 (#44791)
- 15分鐘後再回來>"<

先看一下 QA 再繼續 ~

A man in a red hoodie and light-colored pants is longboarding down a paved road. He is leaning forward in a crouched position, looking back over his shoulder. The road curves to the right, leading towards a range of green, rolling hills under a bright, slightly cloudy sky.

# 傑華老師的煩惱

- 院長突然約傑華吃午餐？
- Python 簡介
- 網路爬蟲實作
- 資料分析與視覺化

# 為什麼院長突然約傑華吃午餐？

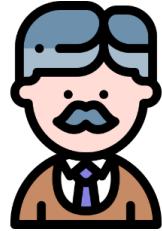
## 就是發生在上周的事情



院長



傑華



在我們人社院裡面，你們社會系的量化研究最厲害，  
而且你們也有請校友來教大數據的爬蟲與分析技術，對嗎？



對的，這都是系上老師和學生一起努力的成果！  
(OS: 糟糕，不知道又要被交辦什麼工作qq)



太好了，我這邊剛好有一份報告想要請你幫忙。  
最近大數據很紅，教育部請我們評估學校在這方面有發展機會嗎？  
預計一個月之後就要，你可以幫這個忙嗎？



…好的，那我來想辦法。  
(OS: 這不是昨天校長才交給你的任務嗎...)



院長



傑華



(回到辦公室後…)



每次有麻煩的事情就丟給我，  
而且時間又這麼趕，該怎麼辦比較好…



對了！  
記得去年網路爬蟲工作坊有教怎麼用 Python 來爬資料和做分析！

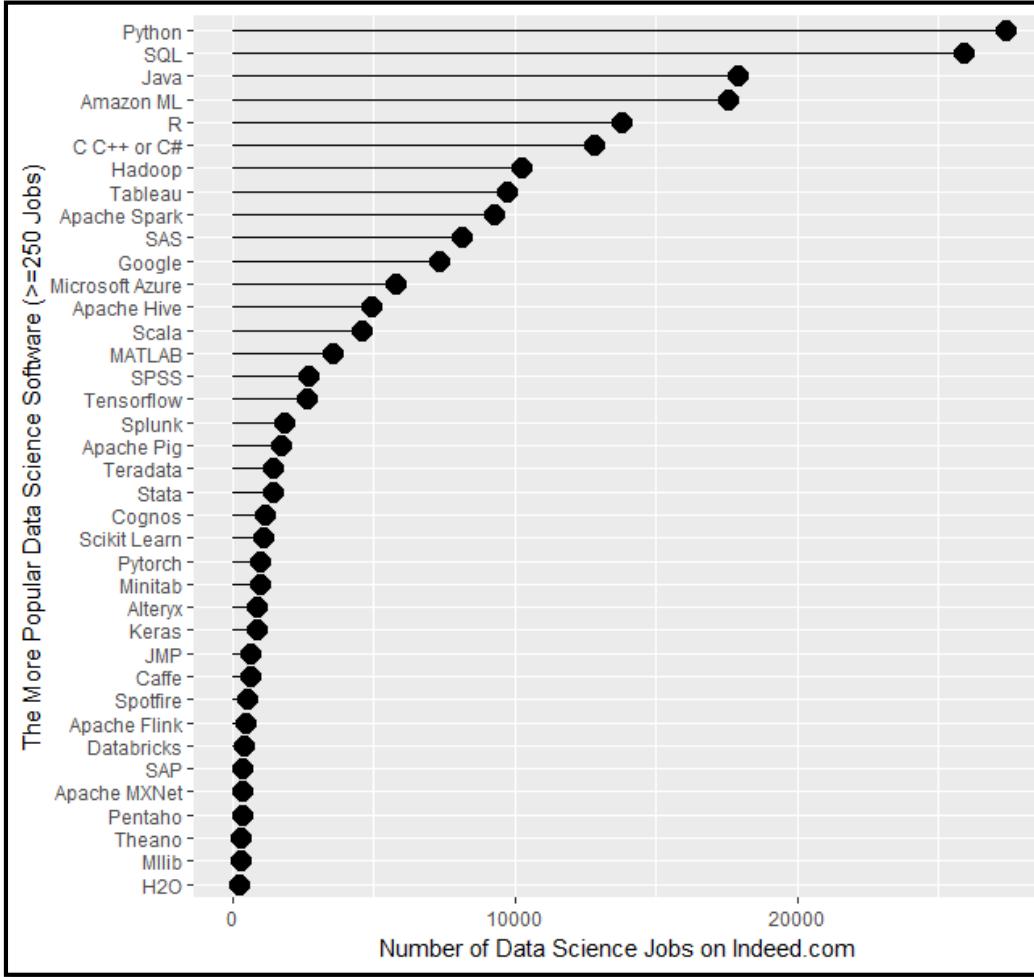


來找學生們幫忙看看好了！

# Python 簡介

# Python簡介

## 分析工具這麼多，為什麼是 Python



- 會 SPSS 還需要 Python 嗎？
  - 一次性 v.s. 重複性
  - 封閉 v.s. 開源
  - 單建模 v.s. 全分析流程
  - ...

稍後我們也可以看一下104的職缺，  
要求求職者會哪些技能

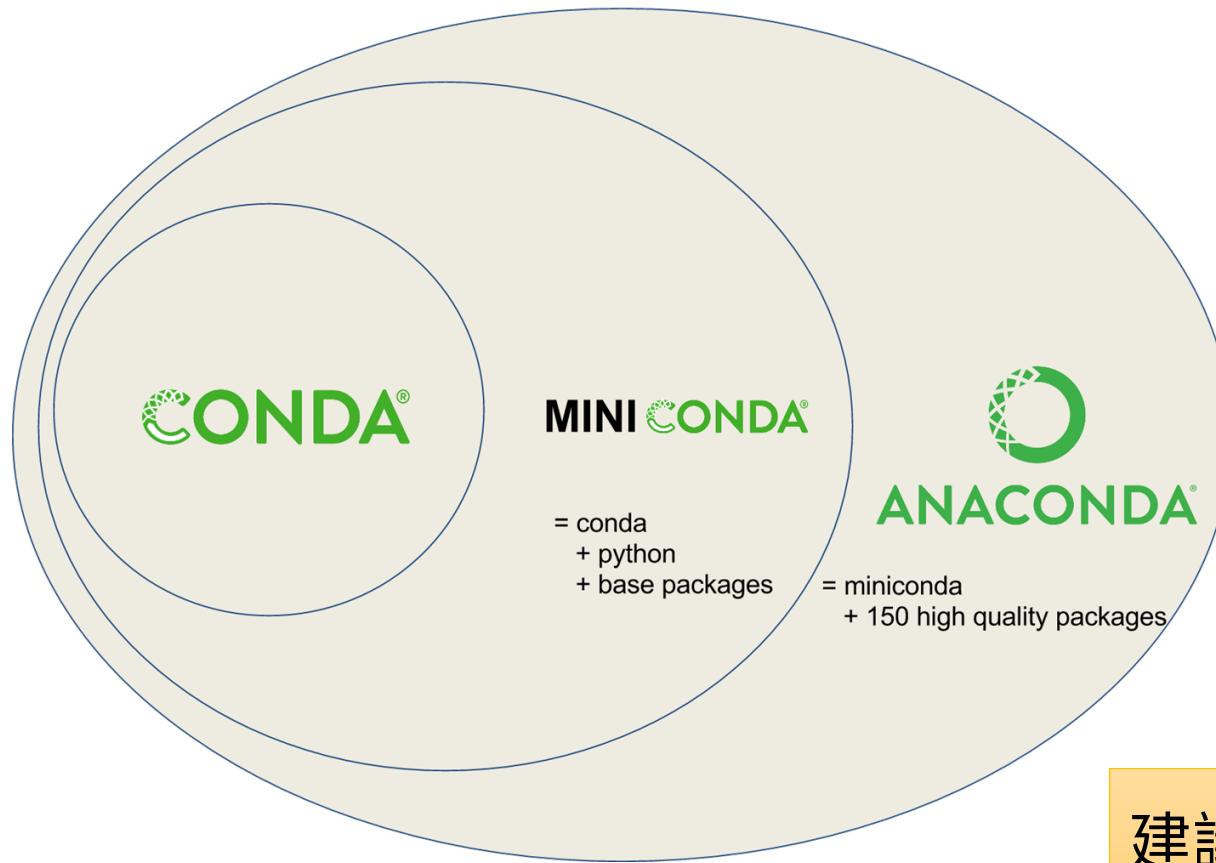
# Python 簡介

關於社會學

數據分析

傑華老師的煩惱

總結



- Anaconda 把所有資料科學專案中常用的套件都預先幫你安裝好，但相對也有檔案肥大的問題！

建議透過 MiniConda 優雅的用python!

Ref: [Getting Started with Conda](#)

科系、職務與待遇：大數據時代下的社會學與勞動市場

Mail: TLYu0419@gmail.com (#44791)

# 安裝Python

## Python 下載地址

[關於社會學](#)[數據分析](#)[傑華老師的煩惱](#)[總結](#)

- <https://docs.conda.io/en/latest/miniconda.html>

### Miniconda

Miniconda is a free minimal installer for conda. It is a small, bootstrap version of Anaconda that includes only conda, Python, the packages they depend on, and a small number of other useful packages, including pip, zlib and a few others. Use the [conda install command](#) to install 720+ additional conda packages from the Anaconda repository.

[See if Miniconda is right for you.](#)

#### Windows installers

Windows

Python version	Name	Size	SHA256 hash
Python 3.8	Miniconda3 Windows 64-bit	55.7 MiB	1f4fff67f051c815b6008f144fdc4c3092af2805301d248b56281c36c1f4333e5
	Miniconda3 Windows 32-bit	49.6 MiB	415920293ae005a17afaef4c275bd910b06c07d8adf5e0cbc9c69f0f890df976
Python 2.7	Miniconda2 Windows 64-bit	54.1 MiB	6973025404832944e074bf02bda8c4594980eed4707bb51baa8fdbba4bf326c
	Miniconda2 Windows 32-bit	47.7 MiB	c8049d26f8b6b954b57bcd4e99ad72d1ffa13f4a6b218e64e641504437b2617b

# 安裝 Python

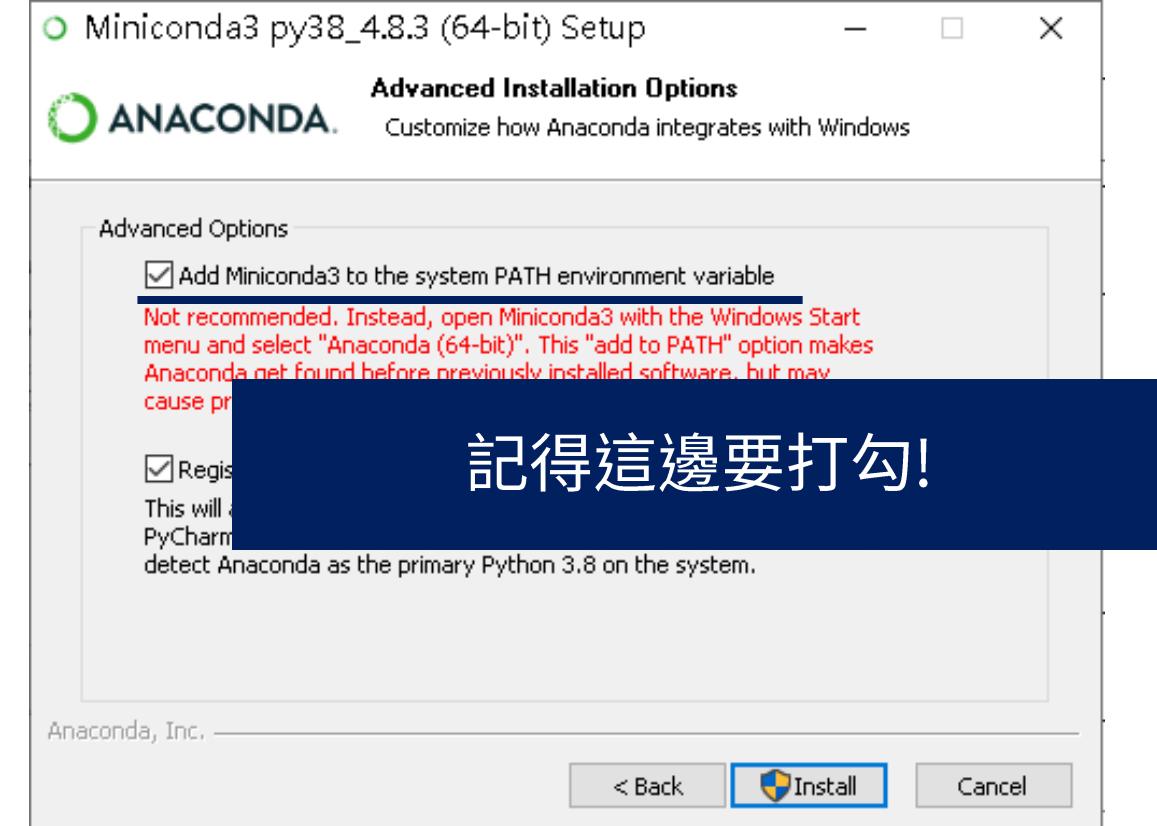
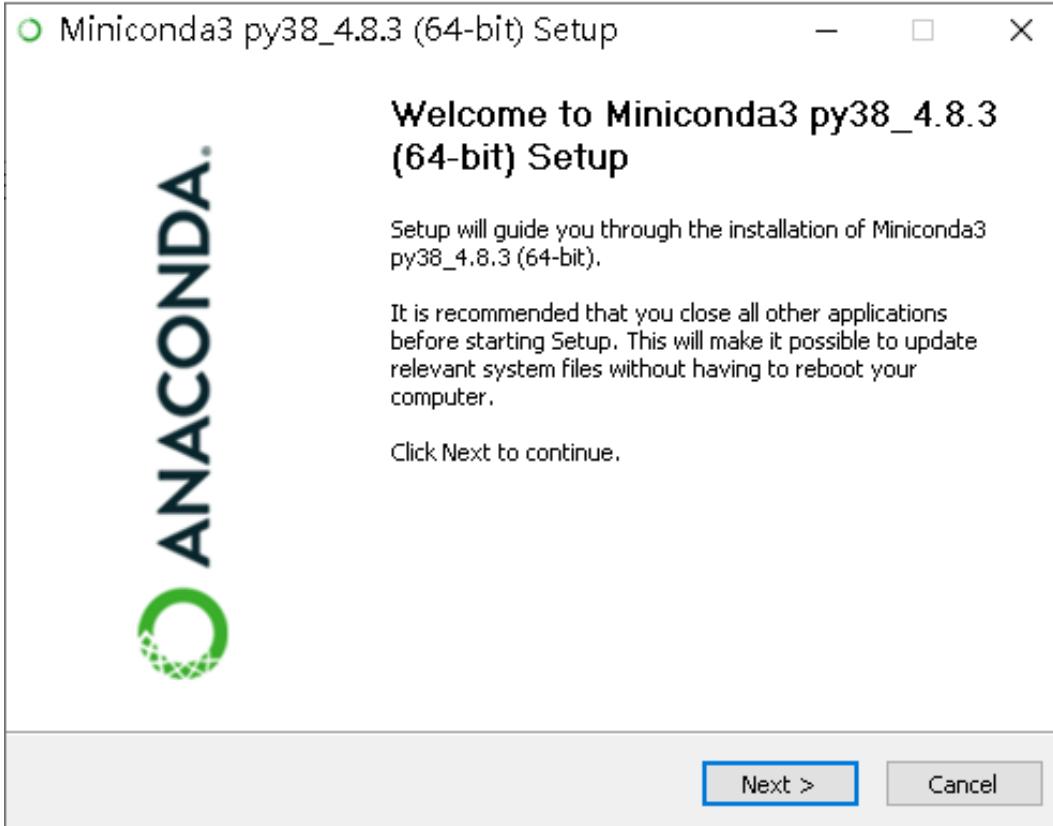
記得環境變數的設定要打勾!

關於社會學

數據分析

傑華老師的煩惱

總結



# 開始實做之前要做的事

# 先別急著動手做

比起怎麼做更重要的是為什麼要做

關於社會學

數據分析

傑華老師的煩惱

總結

- 我只是想要湊學習時數，一定要學這個嗎>"<
- 這個還可以幫你賺一些零用錢，可以認真考慮看看XD



# 人力銀行爬蟲實做

# 啟動 Jupyter

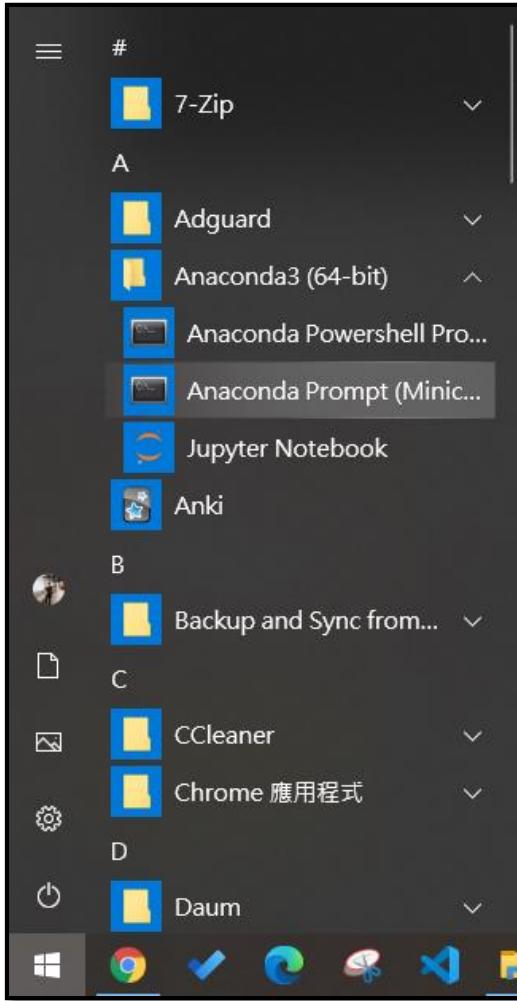
jupyter 是 python 的使用者介面

關於社會學

數據分析

傑華老師的煩惱

總結



系統管理員: Anaconda Prompt (Miniconda3)

(base) C:\Users\TL\_Yu>jupyter notebook

輸入 jupyter notebook 後  
就可以開始使用python囉!

A screenshot of the Anaconda Prompt window. The title bar says '系統管理員: Anaconda Prompt (Miniconda3)'. The command '(base) C:\Users\TL\_Yu>jupyter notebook' is typed into the prompt. A yellow callout box with black text is overlaid on the right side of the screen, containing the text '輸入 jupyter notebook 後' and '就可以開始使用python囉!'.

去年講過開發爬蟲程式的流程  
有人還記得嗎XD

# 人力銀行網路爬蟲實作

## 確認網站結構

[關於社會學](#)[數據分析](#)[傑華老師的煩惱](#)[總結](#)

The screenshot shows the homepage of 104人力銀行 (104 Job Bank). At the top, there's a navigation bar with links like '104人力銀行', 'My104', and various job categories. Below the header, there's a search bar with fields for '關鍵字' (Keywords), '地區' (Region), '職務類別' (Job Type), and a large orange '搜尋' (Search) button. To the left, there's a sidebar with a silhouette of a person looking through a telescope, followed by a list of job search filters such as '金融業急徵', '無經驗可', '月薪4萬元起', etc. On the right side, there are links for '找公司', '公司評論', and '企業品牌大賞'. The main content area features a large banner for '服務業轉職怎麼做?' (How to transition from the service industry?) featuring a woman's portrait. Another banner to the right promotes '全民享補助 美日雙語輕鬆講!' (Everyone enjoys subsidies, English-Japanese bilingual easy-to-understand) with a crowd illustration.

- 透過關鍵字、地區、職務類別來搜尋資料

# 人力銀行網路爬蟲實作

## 確認網站結構

[關於社會學](#)[數據分析](#)[傑華老師的煩惱](#)[總結](#)

104.com.tw/jobs/search/?area=6001001000&jobsource=2018indexpoc&ro=0

104人力銀行 My104 家教 外包 獵才 掌聲 更多 科技大廠聯合招募 → 搜尋薪優工作

關鍵字(例：職稱、公司名、技能專長...) | 台北市

搜尋

只搜尋職務名稱 切換高階職類

更新日期 上班時段 薪資待遇 經歷要求 公司相關 更多條件 排除條件 清空條件 檢視或訂閱條件

全部(74222) 全職(61526) 兼職(10611) 高階(219) 其他 第 1 / 150 頁 日期排序 強力放送

想離職創業？先看完他兩億  
創業生財與故事兩津中

**CCPL BDO - 台北信貸業務【挑戰高獎金的工作在這裡!!】**

渣打國際商業銀行股份有限公司

台北市松山區 | 1年以上 | 專科

【挑戰月入6位數的工作在這裡!!】業務工作不輕鬆，CP值太低怎麼行？渣打信貸BDO適合企圖心強，想靠自己打拼累積財富的您加入！工作內容：•開發新客戶，推廣銷售個人金融商品，與達成各產品指定業務目標。•視客戶需求洽談貸款申辦事宜

月薪30,000~100,000元 員工3000人 距捷運南京復興站410公尺

**【5:30下班，不用加班】電話諮詢顧問**

- 搜尋後可以看到職缺名稱、公司、縣市區域、工作經驗、科系、職務內容、待遇、應徵人數...等等資訊

留意設定不同的條件後對於網址的影響，  
這是學會爬蟲的關鍵！

# 練習題

拆解以下的網址是什麼意思

關於社會學

數據分析

傑華老師的煩惱

總結

- [https://www.104.com.tw/jobs/search/?ro=2&isnew=30&ar  
ea=6001001007&order=4&asc=1&page=15&mode=  
l&jobsourc...](https://www.104.com.tw/jobs/search/?ro=2&isnew=30&area=6001001007&order=4&asc=1&page=15&mode=l&jobsourc...)

# 練習題

拆解以下的網址是什麼意思

- <https://www.104.com.tw/jobs/search/?ro=2&isnew=30&area=6001001007&order=4&asc=1&page=15&mode=l&jobsou>



# Mode 要怎麼選 摘要模式還是列表模式

[關於社會學](#)[數據分析](#)[傑華老師的煩惱](#)[總結](#)

## 摘要模式

全部(1939) 全職(1871) 兼職(46) 高階(11) 其他 ▾ 第 1 / 97 頁 符合度排序

**駐外財務長(工作地點:東南亞)**  
嘉豐海洋國際股份有限公司  
東南亞 | 5年以上 | 大學  
1.擔任海外事業單位財務經理 2.負責事業單位人事、總務、財務及會計等行政管理。3.協助事業單位營運長經營管理及總公司交辦事項。4.具產業成本計算、結算財報及銀行信貸等相關作業 5.具水產品進出口報關工作經驗佳  
待遇面議 員工2000人

**數位通路研究人員**  
第一銀行\_第一商業銀行股份有限公司  
台北市中正區 | 1年以上 | 專科  
1.國內、外專科(含)以上學校畢業，且已取得畢業證書。2.具資訊系統管理或資料分析經驗 3.擅長獨立思考且具前瞻思維 4.國內外數位通路(含行動裝置)創新發展趨勢研究與相關法令規範資訊蒐集彙整 5.協助網銀、APP等數位通路行為資料分析與客群精  
月薪36,000~55,000元 上市上櫃 員工7290人 距捷運台大醫院站360公尺

**商品開發採購人員**  
異網資訊股份有限公司  
台中市大里區 | 2年以上 | 專科  
我們是經營多年且獲獎無數的親子用品電商品牌-Mallidj親子購物網，現在強力徵求志同道合的夥伴！加入我們可以了解電子商務及行銷策略及如何使用後台並透過大數據分析了解電商的客戶及需求，進而提供給客戶更好的優質服務，我們的工作充滿歡樂及  
月薪30,000~35,000元 員工32人

儲存 應徵

0~5人應徵

30人以上應徵

11~30人應徵

## 列表模式

日期	職務名稱	公司名稱	經歷	學歷	地區	應徵人數
★ 11/05	數位通路研究人員	第一銀行_第一商業銀行股份有限公司	1年	專科	台北市中正區	3
★ 11/05	駐外財務長(工作地點:東南亞)	嘉豐海洋國際股份有限公司	5年	大學	東南亞	3
★ 11/05	(早班)物流理貨/點貨人員(工作地	嘉豐海洋國際股份有限公司	1年	高中	高雄市小港區	3
★ 11/05	稅務規劃 人員	第一銀行_第一商業銀行股份有限公司	3年	大學	台北市中正區	3
★ 11/05	機構洗錢與資恐風險評估人員	第一銀行_第一商業銀行股份有限公司	3年	大學	台北市中正區	3
★ 11/05	資料庫(資深)工程師	富欣發網路科技有限公司	3年	專科	台北市松山區	3
★ 11/05	商品開發採購人員	異網資訊股份有限公司	2年	專科	台中市大里區	3
★ 11/03	資料分析師 (Data Analyst)	亞洲指標數位行銷顧問股份有限公司	不拘	碩士	台北市松山區	3
★ 11/03	數據/資料分析師	藍新科技股份有限公司	2年	大學	台北市南港區	3

摘要模式每個分頁有20筆資料，  
列表模式每個分頁有30筆資料，  
選擇列表模式可以減少發送請求的次數！

# 人力銀行網路爬蟲實作

## 確認網站結構

[關於社會學](#)[數據分析](#)[傑華老師的煩惱](#)[總結](#)

The screenshot shows a web browser displaying the 104.com.tw job search page. The URL in the address bar is 104.com.tw/jobs/search/?ro=0&area=6001001000&order=11&asc=0&p... . The page header includes links for 104人力銀行, My104, 家教, 外包, 獵才, 掌聲, 更多, 科技大廠聯合招募, 搜尋薪資, 帳號, 註冊, 登入.

The main search form has fields for 開關鍵字 (關鍵字例：職稱、公司名、技能專長...) and 台北市 (職務類別). Below the form are filters for 更新日期, 上班時段, 薪資待遇, 經歷要求, 公司相關, 更多條件, and 排除條件. There are also buttons for 願仔 (願仔) and 應徵 (應徵).

The first listing is for 博訊生物科技股份有限公司 | 生化科技研發業, located in 台北市內湖區, 經歷不拘, 學歷不拘. The job description includes requirements like 1.英文中等, 2.懂office操作與製作ppt, 3.分類、審核文件資料, 並輸入建檔, 4.進行資料處理、轉換及整合工作, 5.電話負責接聽、過濾、轉接來電, 並提供資訊及留言服務, 6.維護茶水區、接待區及會議室整潔及管理 (定期有打掃外包人員負責大掃). It offers 月薪25,000~30,000元 and has 20 employees. A callout box highlights the "新版介紹" (New Version Introduction) button.

The second listing is for 富士康物業管理顧問股份有限公司 | 保全樓管相關業, located in 台北市內湖區, 經歷不拘, 學歷不拘. The job description includes managing guests at exhibition sites, vehicle entry control, reporting site conditions, and executing tasks. It offers 月薪41,500元 and has 1000 employees. A callout box highlights the "儲存" (Save) button.

A modal dialog box at the bottom right asks "有新工作立即通知?" (Notify me of new jobs?).

- 往下滑動會自動加載更多的職缺資料
- 但是加載到一定數量後，會需要手動點擊下一頁

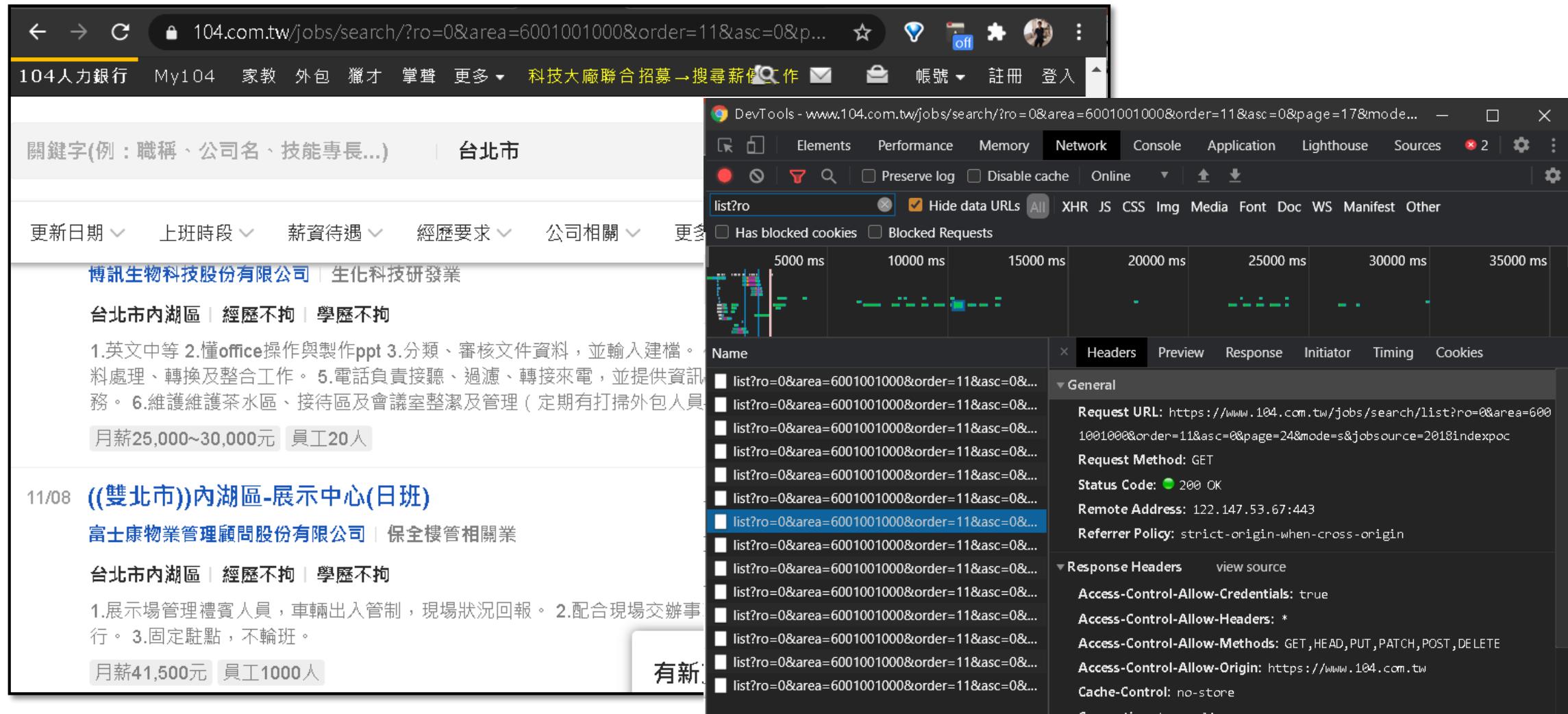
# 人力銀行網路爬蟲實作 透過F12監聽網頁互動方式

# 關於社會學

數據分析

傑華老師的煩惱

總結



# 人力銀行網路爬蟲實作

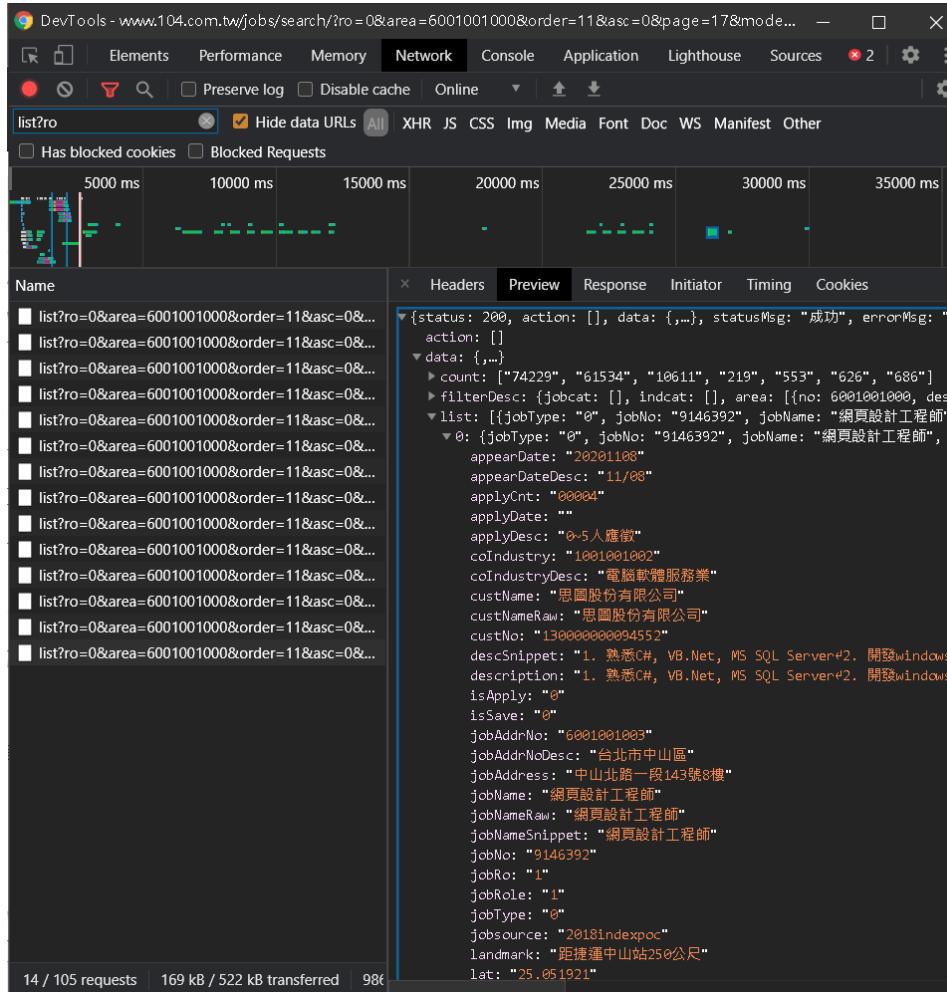
## 利用 ctrl + f 從眾多請求中找到資料

關於社會學

數據分析

傑華老師的煩惱

總結



# 人力銀行網路爬蟲實作

嘗試透過 Python 來取資料

關於社會學

數據分析

傑華老師的煩惱

總結

- Import requests
- From bs4 import BeautifulSoup
- Import pandas

- Pip install xxxxx

# 人力銀行網路爬蟲實作

## 嘗試透過 Python 來取資料

關於社會學

數據分析

傑華老師的煩惱

總結

- url = “ ”
- requests.
- Soup
- pandas

# 網路爬蟲實作

## 迴圈抓更多資料

關於社會學

數據分析

傑華老師的煩惱

總結

# 網路爬蟲實作

完成後趕快把資料存下來>"<

關於社會學

數據分析

傑華老師的煩惱

總結

詳細的職缺資訊還要再往下挖一層  
但時間有限有興趣的再私底下來找我討論>"<

# 休息一下~

- 有問題可以在 Slido 上發問 & 幫別人點讚 (#44791)
- 15分鐘後再回來>"<

先看一下 QA 再繼續 ~

A man in a red hoodie and light-colored pants is longboarding down a paved road. He is leaning forward in a crouched position, looking back over his shoulder. The road curves to the right, leading towards a range of green, rolling hills under a bright, slightly cloudy sky.

# 傑華老師的煩惱

- 院長突然約傑華吃午餐？
- Python 簡介
- 網路爬蟲實作
- 資料分析與視覺化

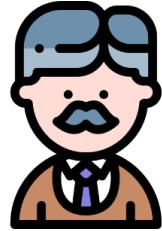
# 一周後 院長又找傑華老師吃飯



院長



傑華



傑華老師，  
上次請你幫忙的工作做得怎麼樣了？



報告院長，我們把台灣最大的人力銀行網站上的職缺都抓下來囉，  
總共有23萬筆職缺的資料呢！(努力邀功一下)



這麼厲害啊！  
那上面會有哪些資料，以及可以做什麼分析呢？



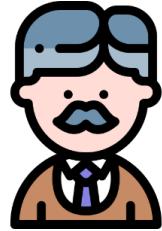
這是雇主在開職缺的時候，對於求職者的期望與勞動狀況，  
可以看到職缺名稱、工作內容、要求技能與待遇等等資訊！



院長



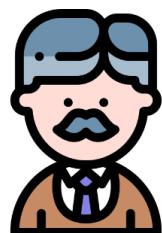
傑華



哇！感覺這真的是很有價值的資料呢！  
相信從這份資料中一定可以找到非常有意義的發現！



沒錯！我們會以社會系為例，  
1. 檢視哪些工作會想要找社會系的學生  
2. 不同科系學生的待遇狀況  
3. 大數據類型的職缺又會要求哪些工具和技能



非常好，我很期待這份分析的資料唷！  
我會跟校長說這份資料是社會系師生的貢獻，  
也會將分析成果再轉給其他系所，  
相信對他們也會是相當有價值的參考！

所以到底要怎麼找到  
高薪的職缺 >”<

# 網路爬蟲實作

所以還是要乖乖腳踏實地的找工作哦 XD

關於社會學

數據分析

傑華老師的煩惱

總結

托嬰中心保母人員 11/08更新

新北市私立甄愛寶貝托嬰中心 本公司其他工作

## 工作內容

- 1.0-3歲嬰幼兒日常托育照顧
- 2.擔任嬰幼兒學習模仿的身教典範
- 3.安排適合嬰幼兒身心發展之各種遊戲活動
- 4.教導嬰幼兒良好的基本生活習慣，並保護其安全

職務類別 幼教班老師、照顧服務員

工作待遇 **日薪25,000~35,000元**

上班地點 新北市中和區景平路240巷1弄10號 ☺

倉管及物流管理 11/06更新

普泰光電股份有限公司 本公司其他工作

## 工作內容

- 1.定期清點物料，並將物料標示並放於正確位置，必要對相關人解說，
- 2.定期清潔庫存、工具及設備或儲藏室，以確保符合安全規章及存貨安
- 3.貨品輸送與成品包裝
- 4.需搬運重物
- 5.不定期需支援產線生產製造

職務類別 倉儲物流人員、倉管

工作待遇 **日薪25,000~32,000元**

上班地點 桃園市中壢區中壢工業區 ☺

# 哪些工作會想要找 社會系的學生？

# 網路爬蟲實作

用文字雲來視覺化分析結果吧!

關於社會學

數據分析

傑華老師的煩惱

總結



# 勞動市場給不同科系學生的 待遇怎麼樣？

# 網路爬蟲實作

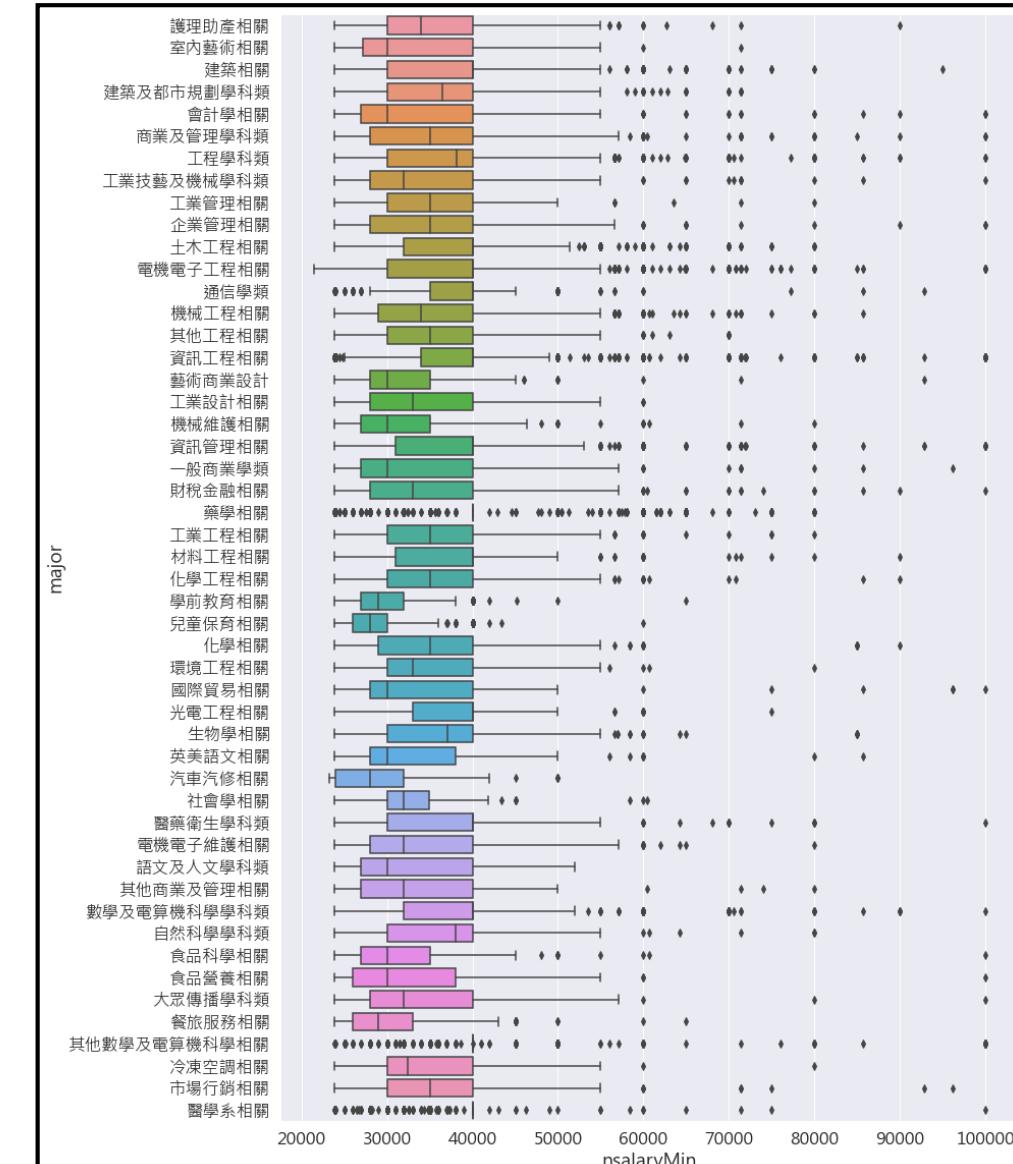
這是常見的箱型圖，你會解讀他嗎XD

關於社會學

數據分析

傑華老師的煩惱

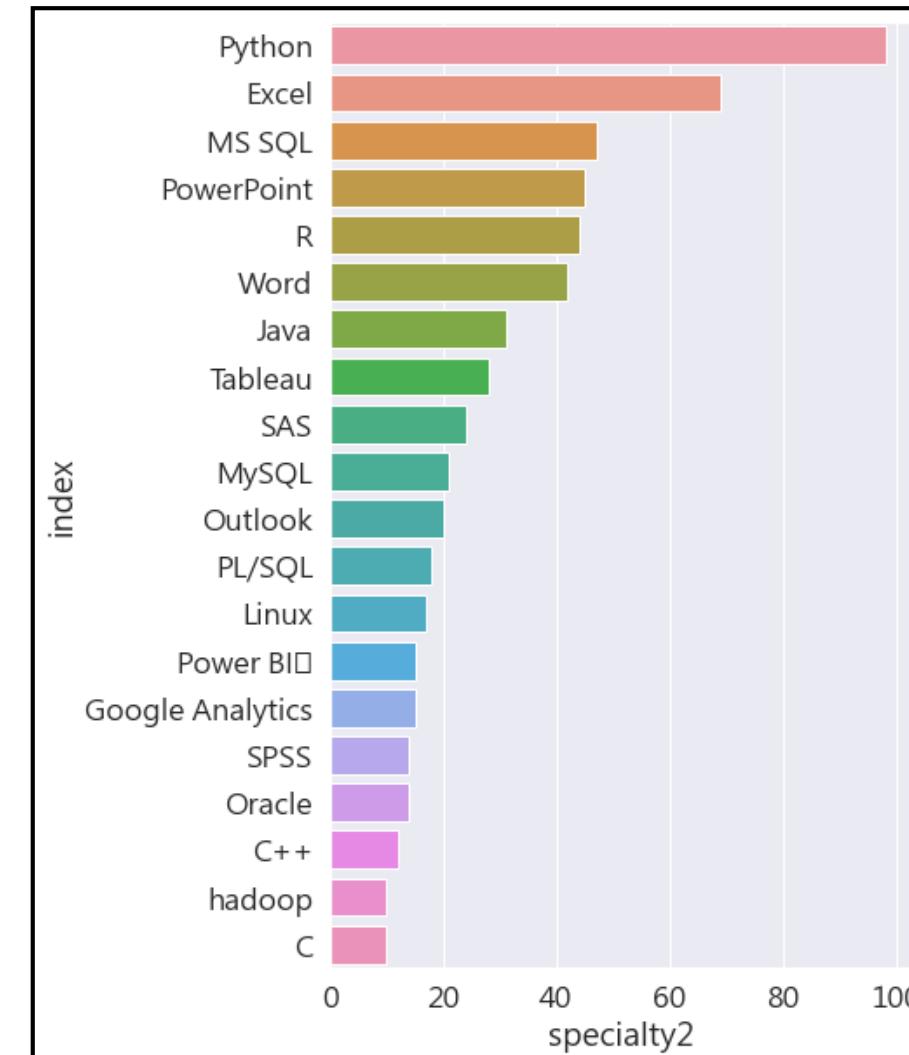
總結



# 資料科學的工作要求那些技能？

# 網路爬蟲實作

## 不意外第一名就是 Python

[關於社會學](#)[數據分析](#)[傑華老師的煩惱](#)[總結](#)

先看一下 QA 再繼續 ~

# 總結

# 確認目標

把問題說清楚，講明白！



## 學校

- 要設法提升能見度，讓企業能看見社會系的特色
- 多蒐集畢業學生的流向，並更積極的給學生職涯建議
- 提升分析的工具與技術
- 簡報設計與製作

關於社會學

數據分析

傑華老師的煩惱

總結



## 學生

- 做一個有深度的專案，勝過做許多蜻蜓點水的專案
- 鼓勵大家從小地方著手，發現問題後嘗試採取行動解決問題！
- 提升分析的工具與技術
- 簡報設計與製作

# 確認目標

把問題說清楚，講明白！



## 學校

- 要設法提升能見度，讓企業能看見社會系的特色
- 多蒐集畢業學生的流向，並更積極的給學生職涯建議
- 提升分析的工具與技術
- 簡報設計與製作

關於社會學

數據分析

傑華老師的煩惱

總結



## 學生

- 做一個有深度的專案，勝過做許多蜻蜓點水的專案
- 鼓勵大家從小地方著手，發現問題後嘗試採取行動解決問題！
- 提升分析的工具與技術

雖然資料科學類的職缺主要是找資訊、統計系的學生，但仍有42%的職缺沒有限制科系背景，建議可以商業分析師作為求職的目標！

如果跟學長一樣反骨  
可以再進一步往資料科學家發展 >“<

- Facebook
  - Python Taiwan  
<https://www.facebook.com/groups/pythontw>
  - 資料科學小嫩嫩: <https://www.facebook.com/groups/774141029405112>
  - <https://www.facebook.com/tlyu0419>
- Udemy
  - <https://www.udemy.com/>
- Youtube
  - 大數軟體 <https://www.youtube.com/channel/UCFdTiwvDjyc62DBWrIYDtIQ>
  - Pydata [https://www.youtube.com/channel/UCOjD18EJYcsBog4lozkF\\_7w](https://www.youtube.com/channel/UCOjD18EJYcsBog4lozkF_7w)
  - PyCon Taiwan <https://www.youtube.com/c/PyConTaiwan/featured>

# 貼心小叮嚀

不是每位老師都喜歡計量研究  
使用前請事先調查老師的口味

- <https://forms.gle/re3R3b5NQxiy7V2E9>
- 我會匿名、整理然後幫忙回饋給老師