

A man in a red hoodie and light-colored pants is longboarding down a winding asphalt road. The background features a vast, green, hilly landscape under a bright, cloudy sky at sunset or sunrise.

科系、職務與待遇： 大數據時代下的社會學與勞動市場

- 畢業校友 - 游騰林
- 202011



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](https://www.slido.com)
#meet20

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？ | • 系上 |
| • 畢業之後呢 | • 反面應用案例 | • Python 簡介 | • 學生 |
| • 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

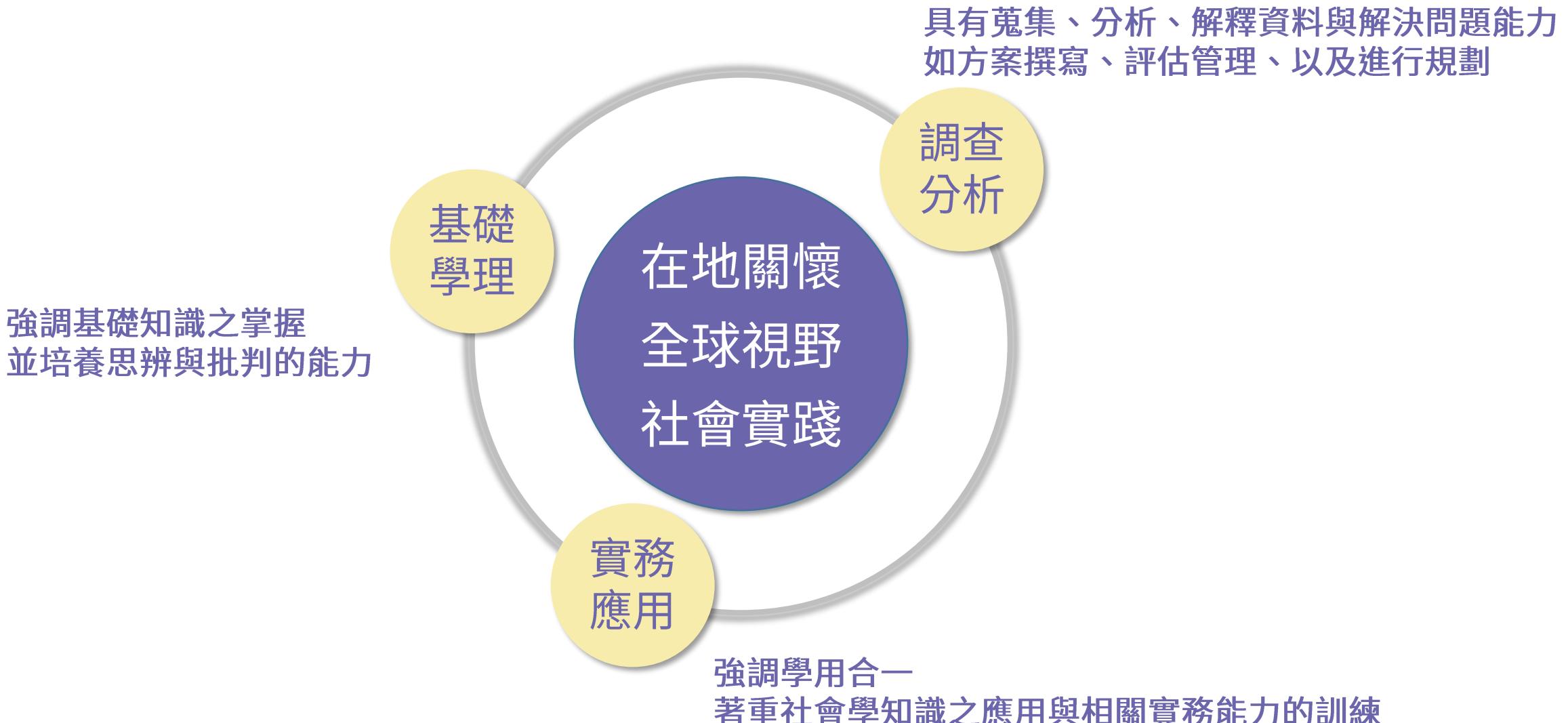
基礎學理、調查分析與實務應用

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

為什麼要教育改革?

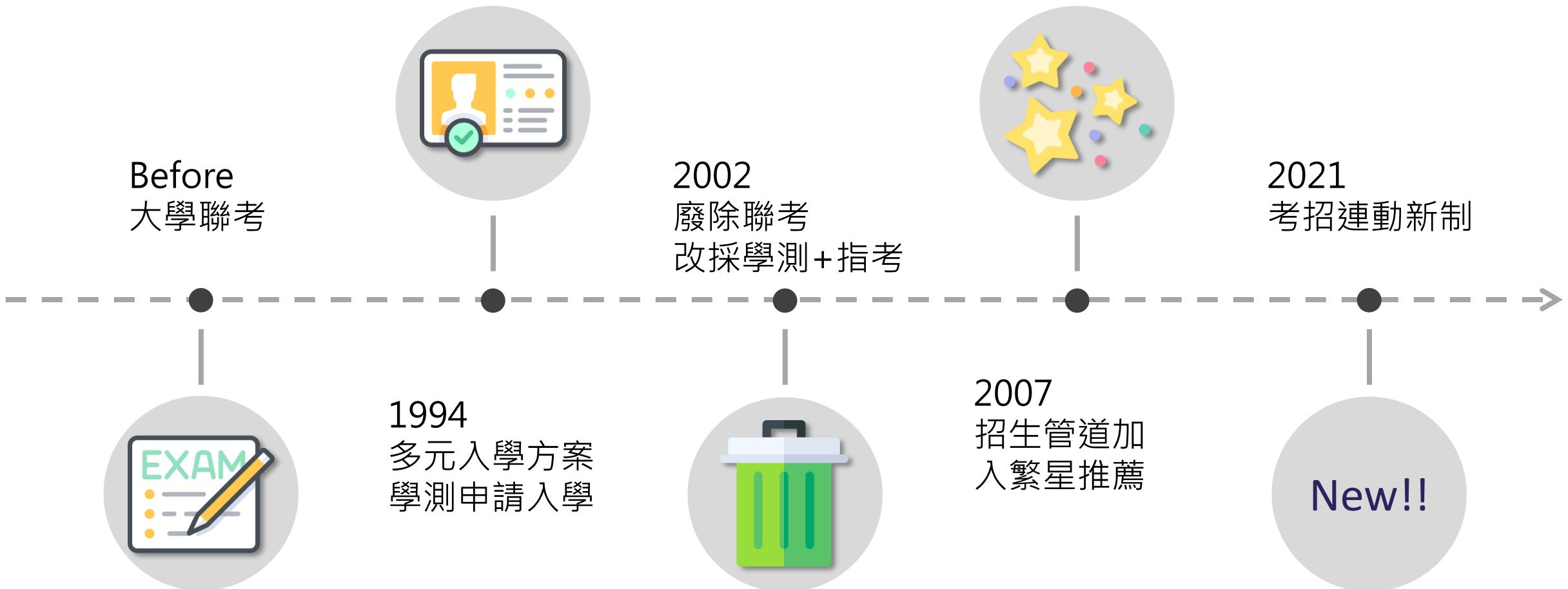
改革是希望教育機會的均等，但也衍生出許多弊端

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

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>

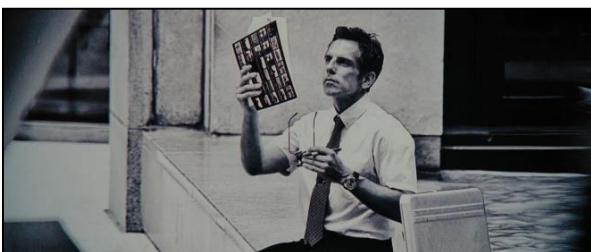
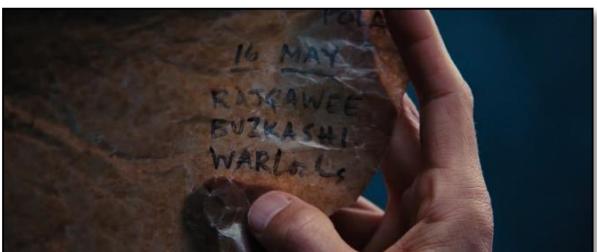
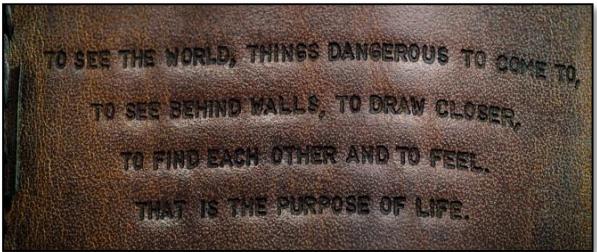
但是畢業之後呢？



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

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-游承輝	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！

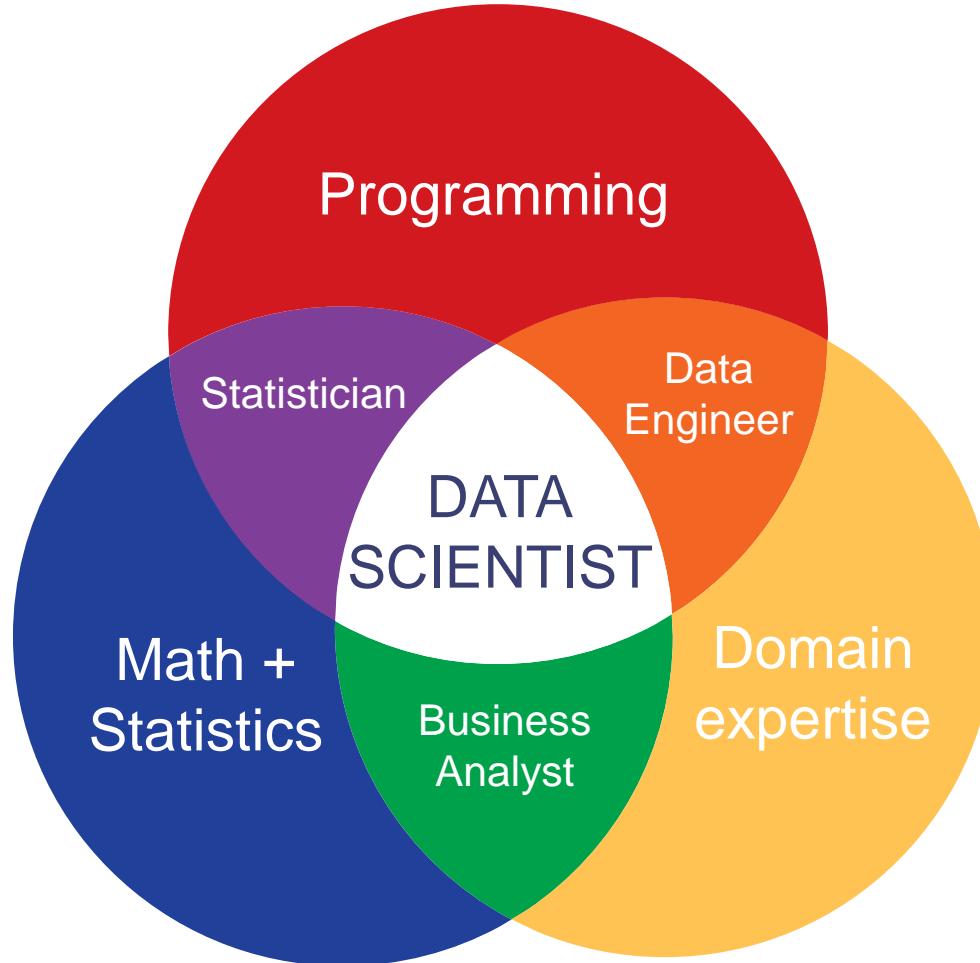
Ref: [社會學系過來人經驗分享](#)

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

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

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

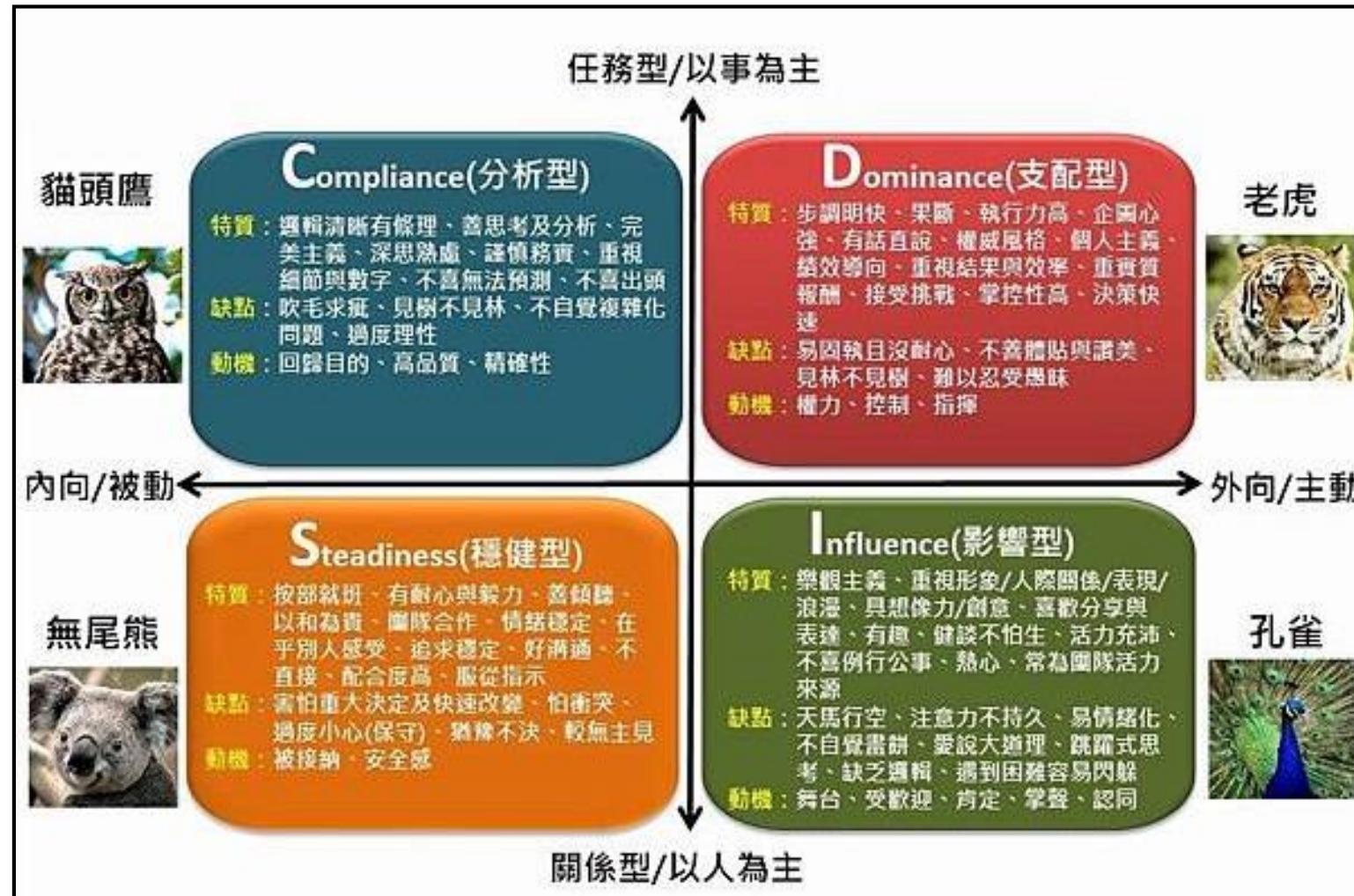
DISC人格特質測驗說明

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
- 10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

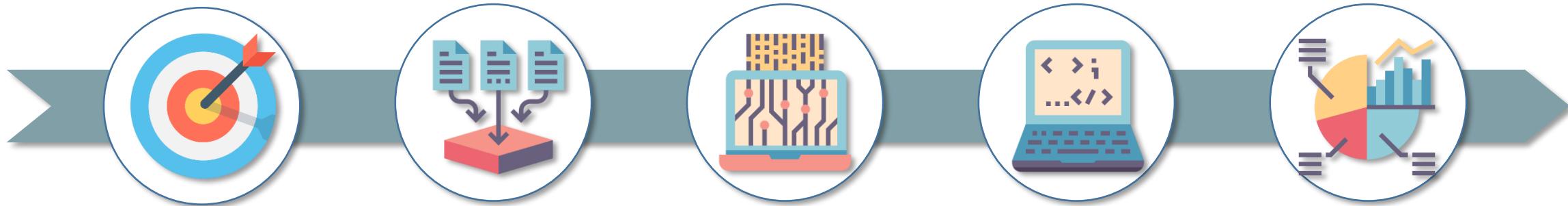
面試時會問，一定要記得很清楚

關於社會學

數據分析

傑華老師的煩惱

總結



確認目標

- 問題重要嗎？
- 問題定義

資料蒐集

- 資料在哪裡?
- 資料品質
- 要怎麼取得資料?

預處理

- 遺漏值
- 離群值
- 標準化
- 特徵工程

模型建置 專案佈署

- 演算法選擇
 - 選定評估指標
 - 調參
 - 不平衡資料
- 撰寫分析報告
 - 專案上線

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](#)

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
- ...

分析流程_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

- 公眾利益 / 影響

[停車方針](#)、[計程車](#)

- 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...

分析流程_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
2019-05-05
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
2019-05-05
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

分析流程_資料蒐集

資料在哪裡？要怎麼取得資料？

關於社會學

數據分析

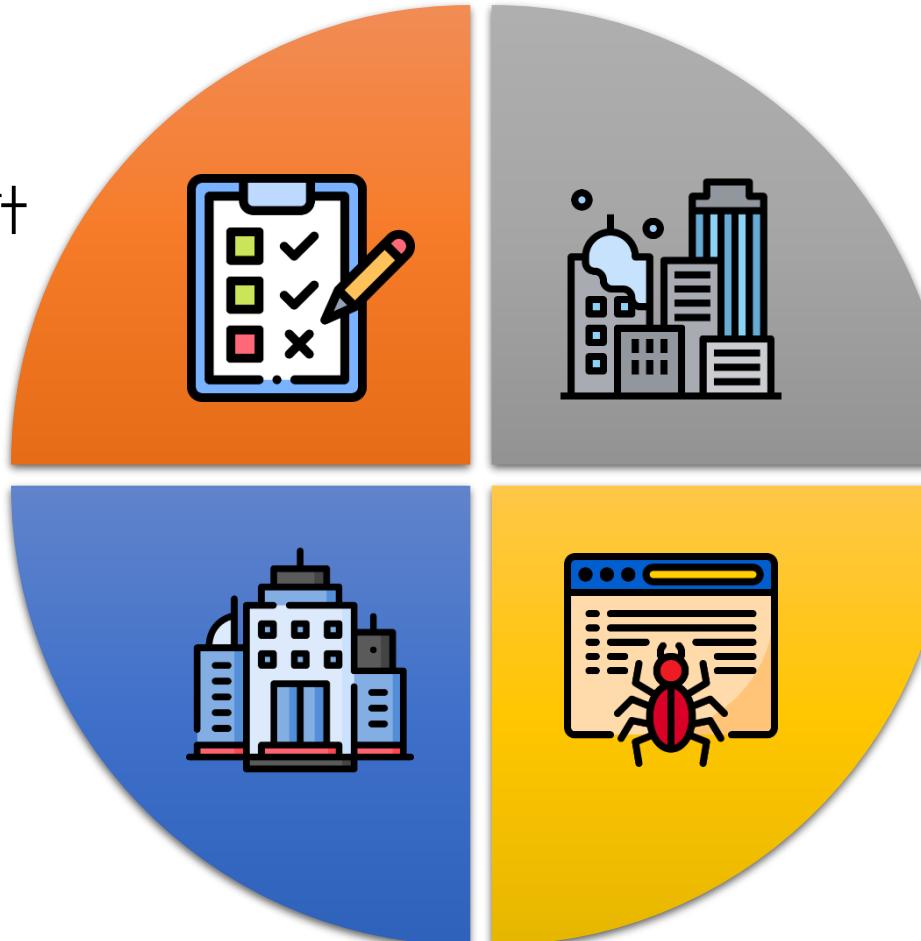
傑華老師的煩惱

總結

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](#)

分析流程_資料蒐集

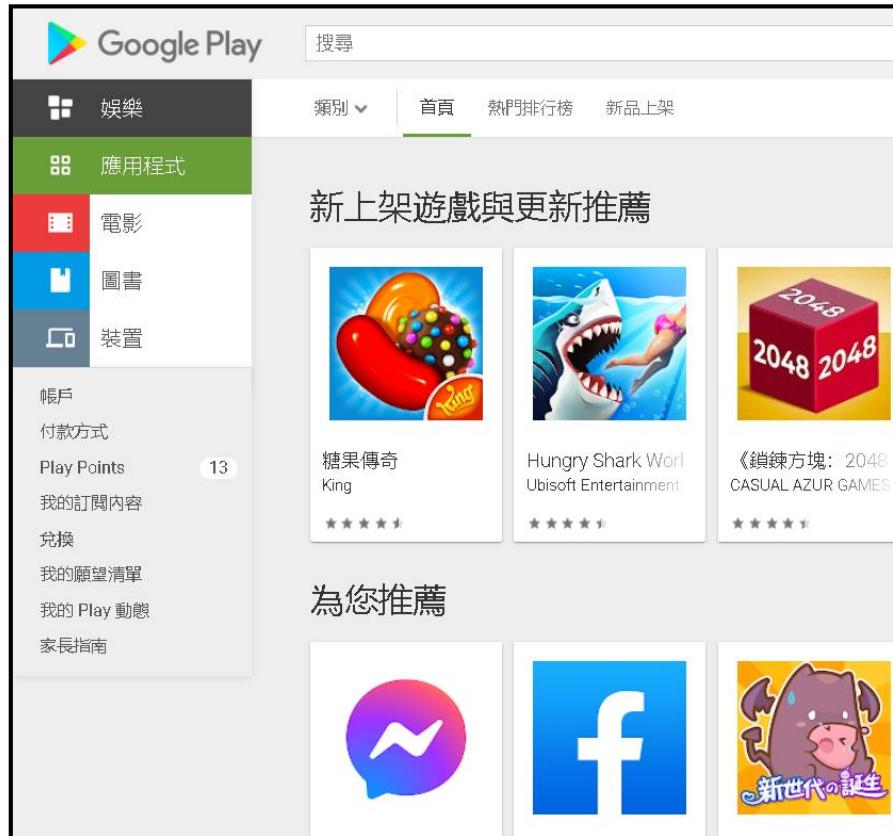
資料在哪裡？要怎麼取得資料？

關於社會學

數據分析

傑華老師的煩惱

總結



APP 名稱

Facebook

★★★★★ 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

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 the text '筆記型電腦'. Below the search bar, there is a large blue banner with the text '商品名稱、價格、滿意度、銷售量' (Product Name, Price, Satisfaction, Sales Volume). On the left side, there is a promotional banner 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' (Free shipping every day, starting from \$11). There are also sections for '條件篩選' (Filter Conditions) with categories like '分類' (Category), '更多' (More), '出貨地點' (Shipment Location), '運送方式' (Delivery Method), and '品牌' (Brand). The main content area displays a grid of laptop products with their names, prices, ratings, and sale status. For example, one listing for an Acer S3 UltraBook is priced at \$5,500 with a 5-star rating and 473 sales. Another listing for an ASUS X205TA is priced at \$4,500 with a 5-star rating and 19 sales.

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

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. The main post discusses the impact of the pandemic on movie theaters and mentions her participation in the Taiwan Golden Horse Film Festival (TGHFF). A large image of Tsai speaking at a press conference is displayed. Below the main post, there is a large, bold text overlay: "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". The post has received 17K likes and 916 comments. The comments section is highlighted with a yellow box, showing several replies from Tsai Ing-wen and other users.

關於社會學

數據分析

傑華老師的煩惱

總結

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

資料預處理

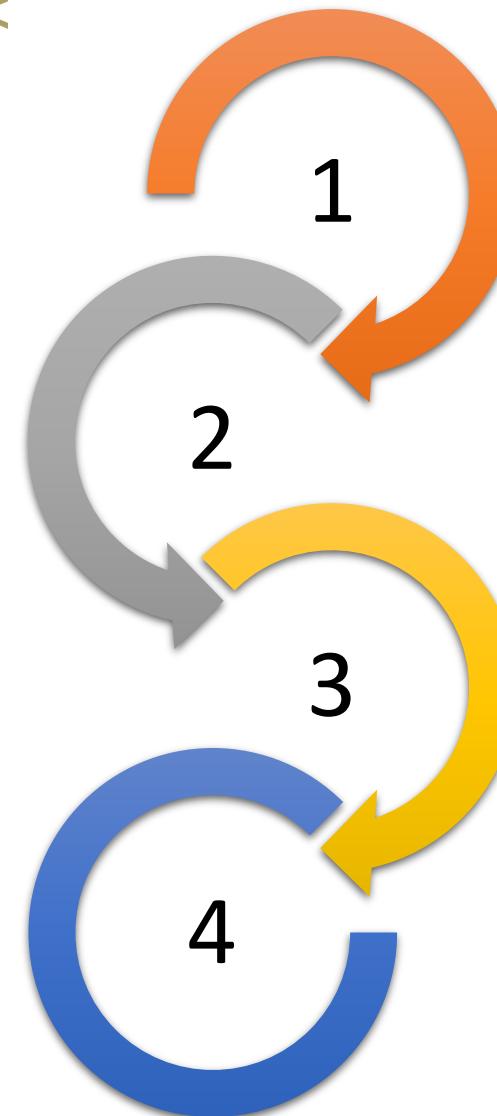
其實這些知識你們都學過>“<

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

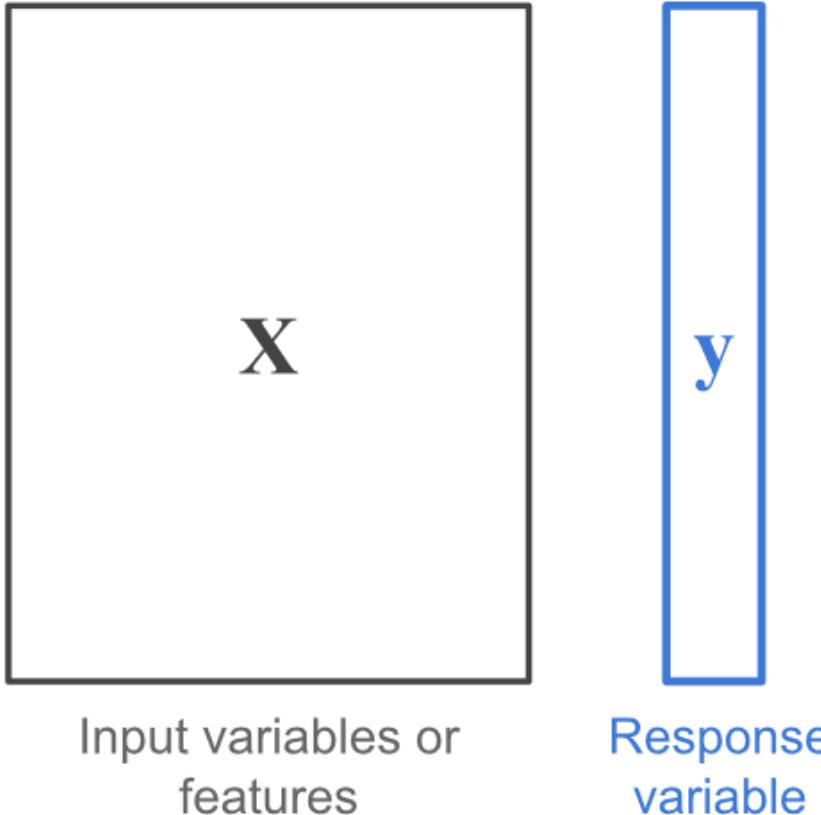
1. Aggregation
2. Normalization
3. Feature selection
4. Discretization
5. Generalization

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

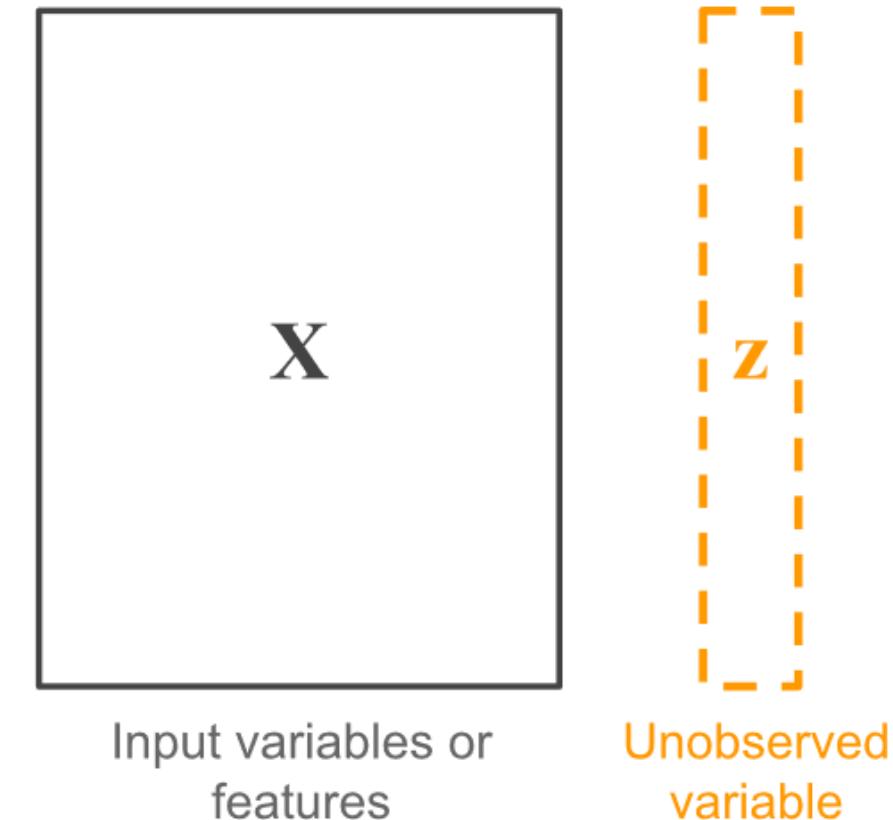
科系、職務與待遇：大數據時代下的社會學與勞動市場

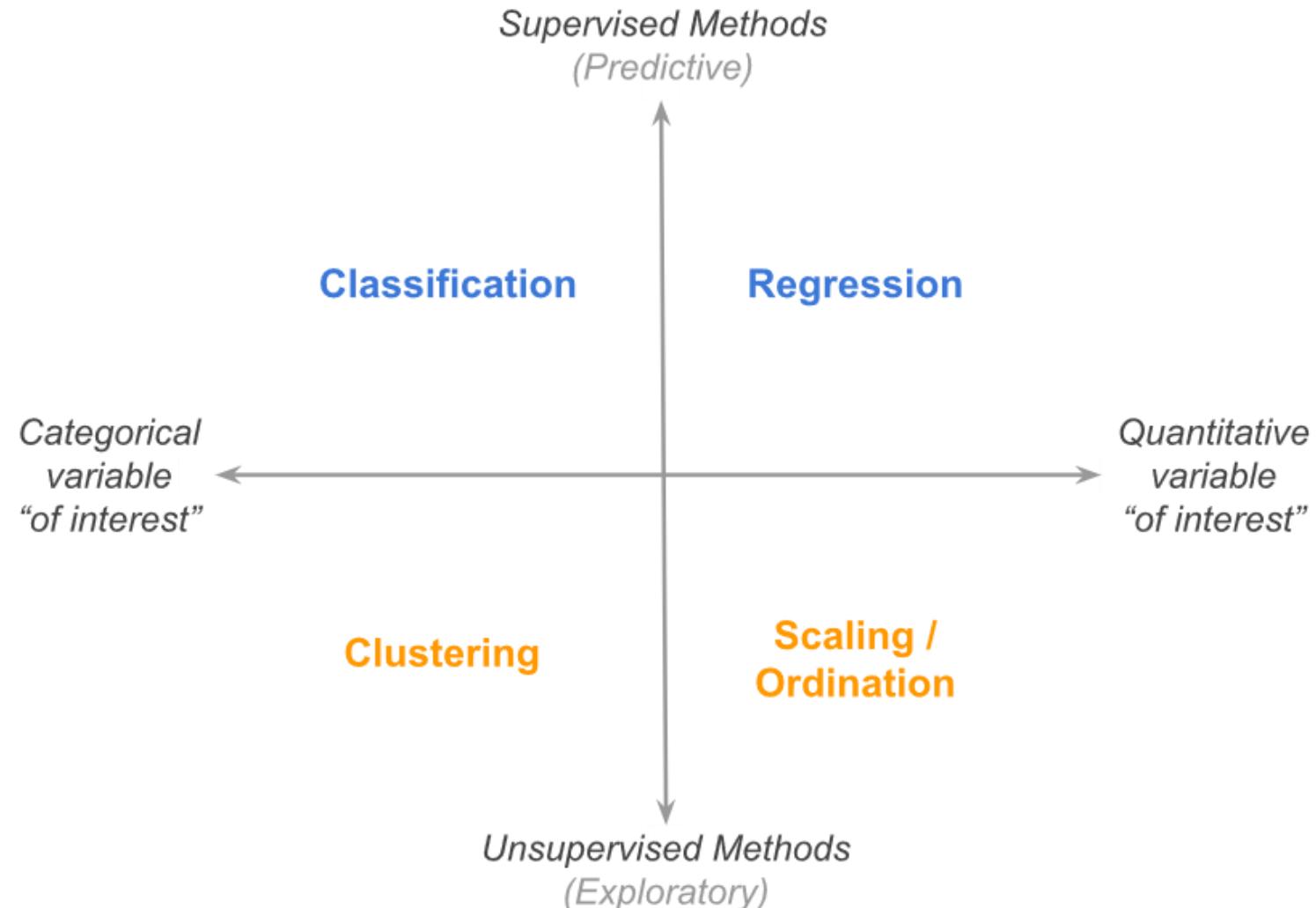
Mail: TLYu0419@gmail.com

Supervised Learning



Unsupervised Learning





- 預測學生延畢
- 預測是否會升學
- 預測是否會找到工作
- ...



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

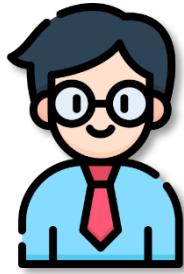
企業怎麼用



- 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

- 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

- 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

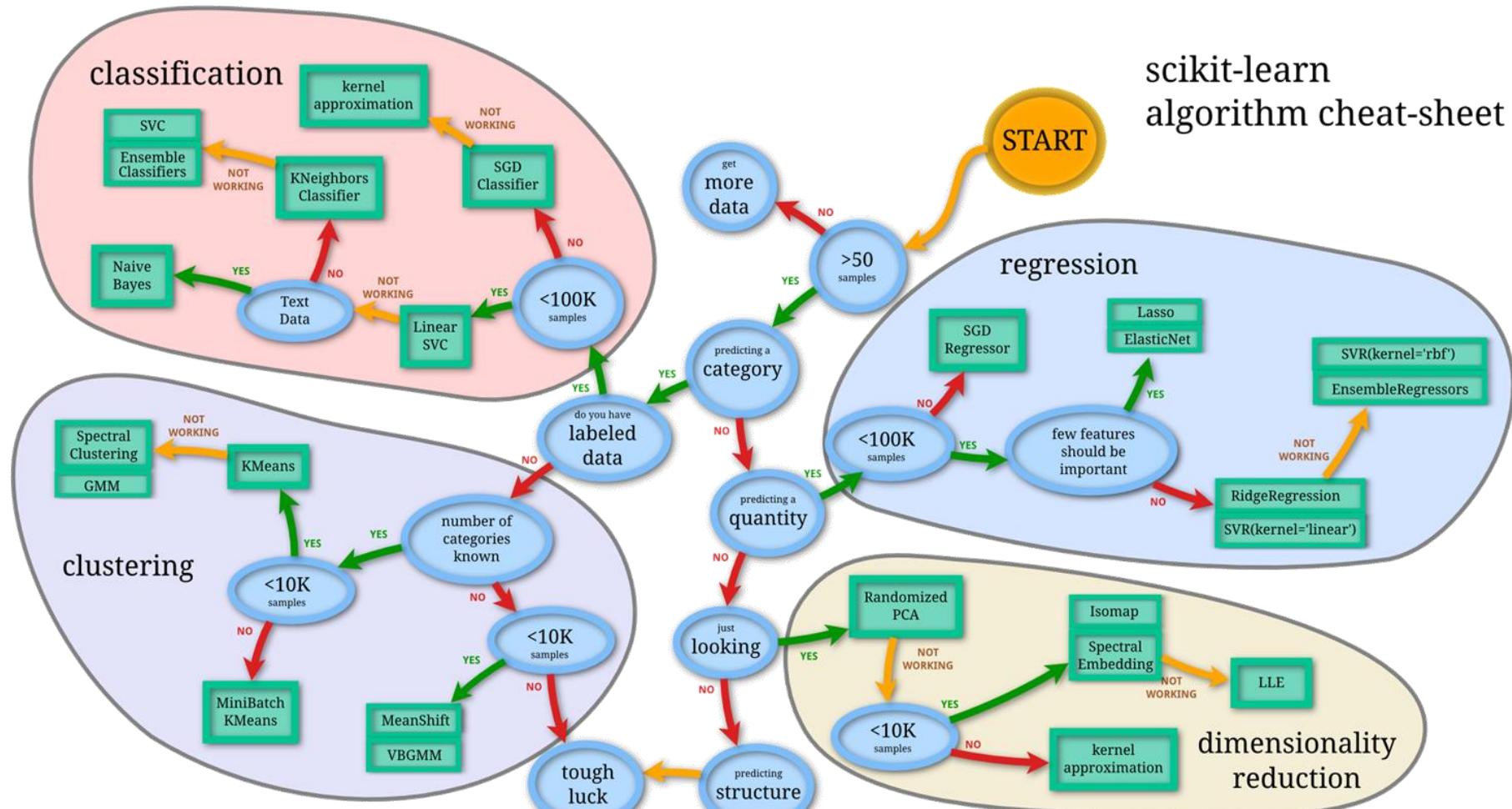
- 欠費行為分群
- 消費行為分群
- 進線行為分群



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

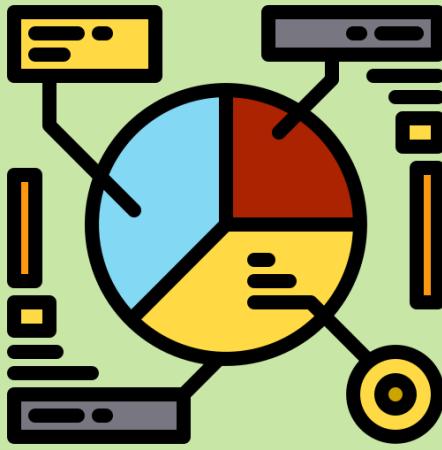
如何選擇合適的演算法



模型部屬

依專案目標規劃後續的應用、上線流程

分析案



- 運用視覺化的技巧，協助長官/讀者輕鬆理解想傳達的理念
- 要有清楚、具體的建議，避免只是說哪裡不平等、不公平

模型案

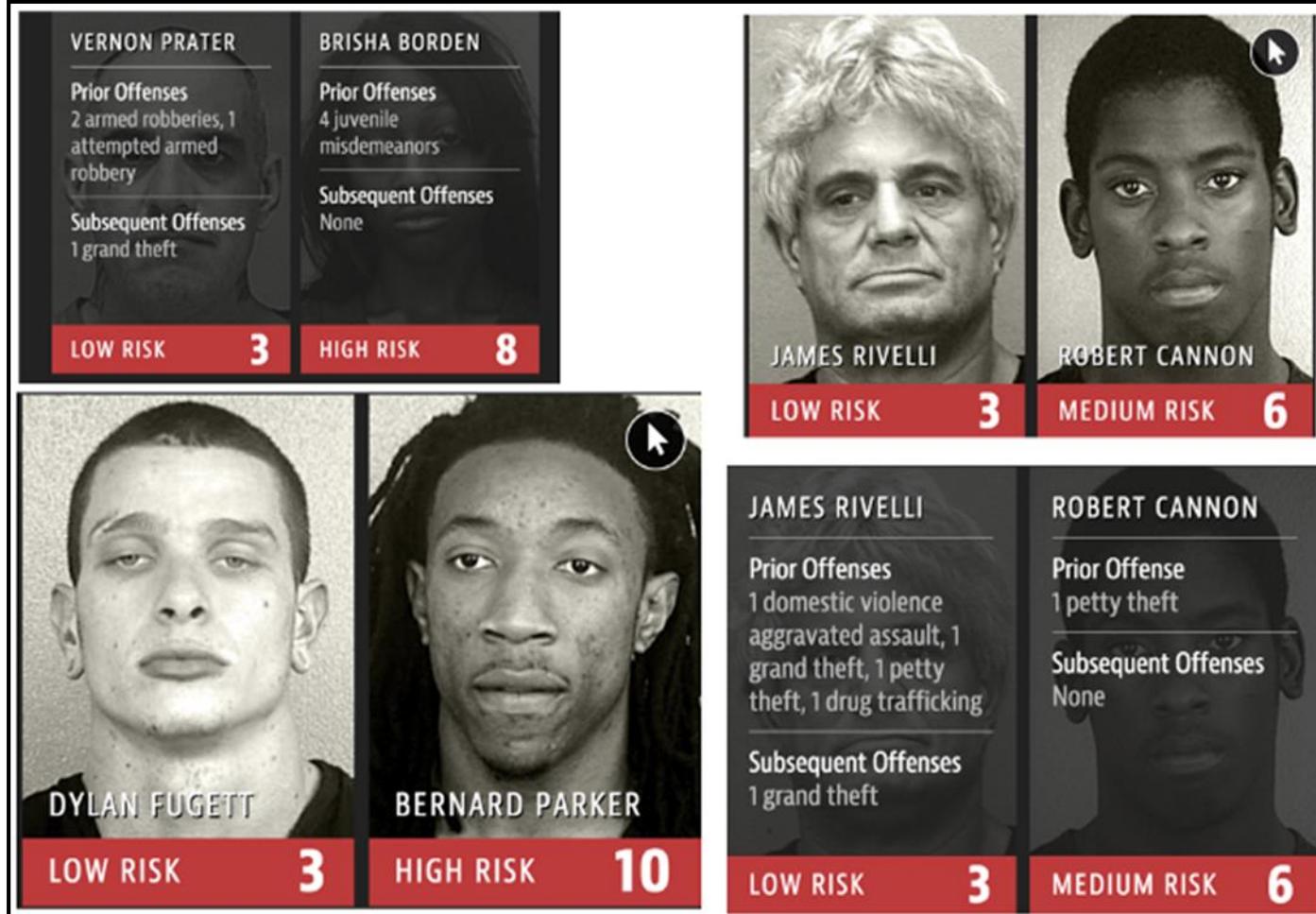


- 透過模型更精準的找到目標客戶、改善業務流程。
- 要清楚的告訴業務單位改變跟影響是什麼？
- 需要設計追蹤的機制定期檢視模型成效

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

在美國，當刑事被告面臨宣判，法官斷最適合的刑罰，這些判斷因素之一「分數」（risk assessment score）是否會犯其他罪行的數字。「風險評估」是根據被告的背景、犯罪歷史、釋金數字、治療計劃或刑期長短。如果「風險評出分數」較高，法官在裁量的時候，會考慮被告未來可能再犯罪的可能性高而科系、職務與待遇：大數據時代下的社會學與勞動市場
Mail: TLYu0419@gmail.com

Ref:1. [Machine Bias](#)

2. [大數據應用潛藏的歧視風險](#)

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

關於社會學

數據分析

傑華老師的煩惱

總結

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

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

The image shows a screenshot of a YouTube interface. In the center is a video player displaying a close-up of two hands tied together with white earphones. To the left of the video player is a sidebar with several video thumbnails and their details:

- 【文曰小強】10分钟速读《三生三世》33万字原著
文曰小强——官方频道
觀看次數：8.6萬次・3年前
- Bruno Mars Jazz & Bossa Nova Cover - Relaxing Caf...
- Cafe Music BGM chan...
觀看次數：178萬次・2年前
- 合輯 - Aakash Gandhi - Heavenly (No Copyright Music)
雅卡什·甘地、Roa、阿里吉特·辛格和其他演出者
觀看次數：50+ (0)
- 世界科幻巨著《索拉里斯星》上集：对宇宙生命的...
幻海航行-science fiction
觀看次數：10:57 (0)

On the right side of the video player, a yellow box highlights a list of interaction options:

- 加入待播清單
- 儲存至「稍後觀看」清單
- 儲存至播放清單
- 不感興趣
- 不要推薦這個頻道** (This option is highlighted in yellow.)
- 檢舉

**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

50

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

IN SIDE 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**

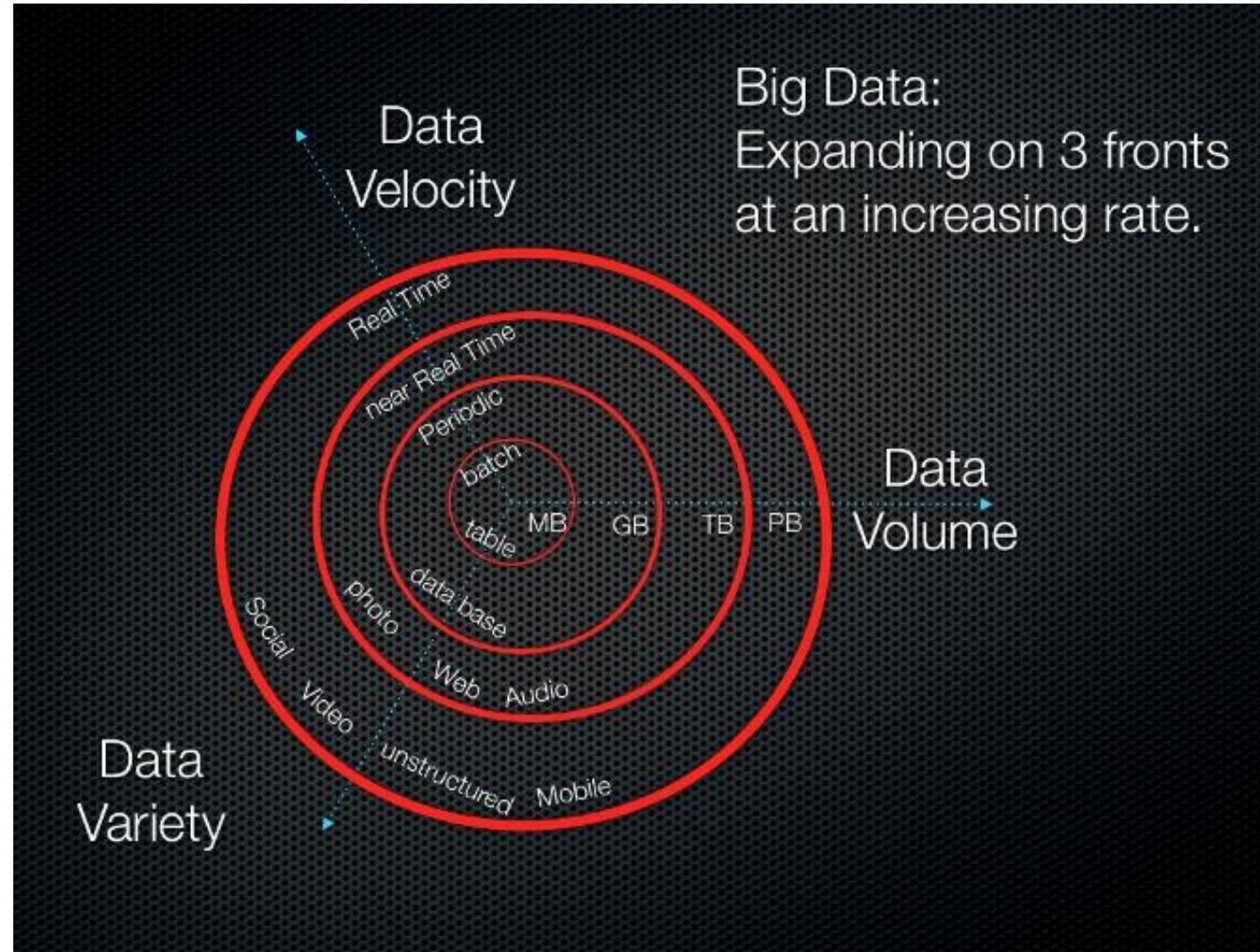
月銷量 1萬

台南市佳里區

到底大數據是好還是壞> “<

大數據給社會學什麼挑戰？

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

從數據角度出發，
設法改善業務流程、提升銷售量

從社會理論出發，
找出社會的不平等並設法改善

關於社會學

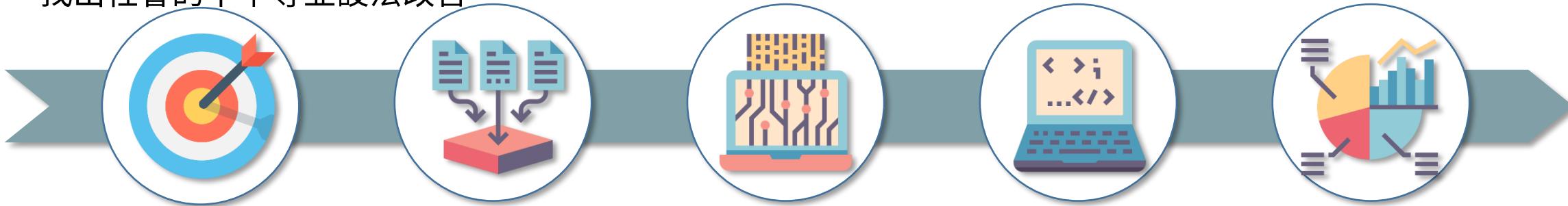
數據分析

傑華老師的煩惱

總結

講究快速解決問題

有嚴謹的研究設計來驗證知識
重視正確性與因果關係



確認目標

資料蒐集

預處理

模型建置 專案佈署

蒐集資料時通常沒有特定的目的，
成本較低，但是資料碎片化問題嚴重

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

機器學習、
深度學習模型

線性回歸、
邏輯回歸

大數據給社會學的挑戰

有請致嘉老師協助回答

關於社會學

數據分析

10月13日(二)

已讀
下午 12:28

老師跟你訪問一下，
老師覺得社會學系的學生

對「時代」意義的掌握和時代中人的行為意義了解/判讀。

若能像你一樣同時掌握技術，就更能有效的預測

下午 12:56

- 挑戰／優勢是什麼
- 社会学曾经的优势有调查数据、有分析工具、有知识积累。这三块是社会学最核心的优势
- 调查数据的话语权变弱是不争的事实。
- 补会学家们剩下的优势只有知识积累了。问题是，知识积累也依靠数据在数据受到冲击的前提下，补会学的知识积累也可能会坐吃山空。归纳起来讲，如果说大数据对补会学研究有什么挑战，其实不是大数据的挑战，而是补会变迁的挑战，我们生活的这个补会变了。
- 们没有能力把握信息社会、数据化社会的特征，社会学就有可能毁在我们手里。因此，让补会学再学习，向补会学习，把握数据化社会的特征，可能是补会学进一步发展的希望。

休息一下~

- 有問題可以在 Slido 上發問 & 幫別人點讚
- 10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

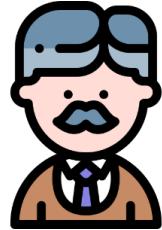
為什麼院長突然約傑華吃午餐？



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

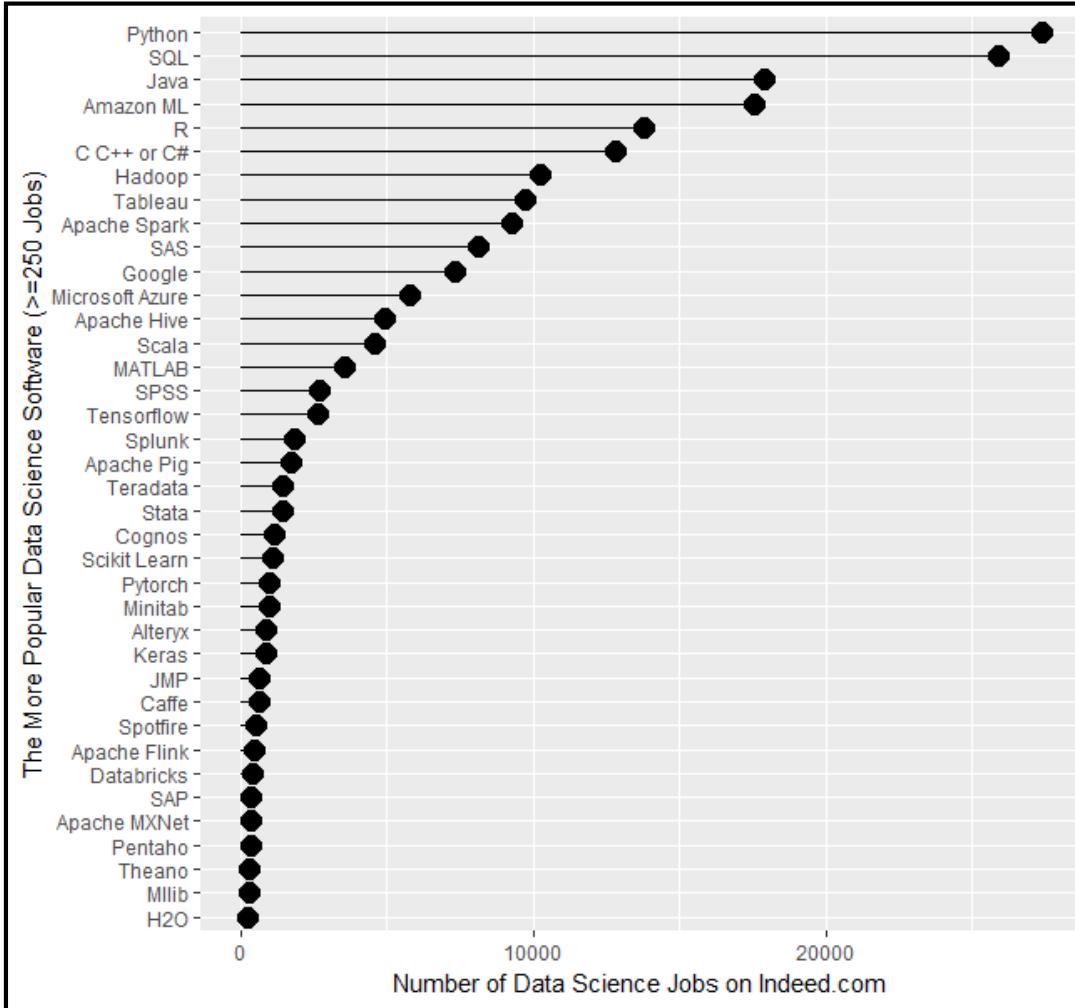
分析工具這麼多，為什麼是 Python

關於社會學

數據分析

傑華老師的煩惱

總結



<http://r4stats.com/articles/popularity/>

- 我會SPSS還需要Python嗎？
- 稍後我們也可以看一下104的職缺，要求技能的數量

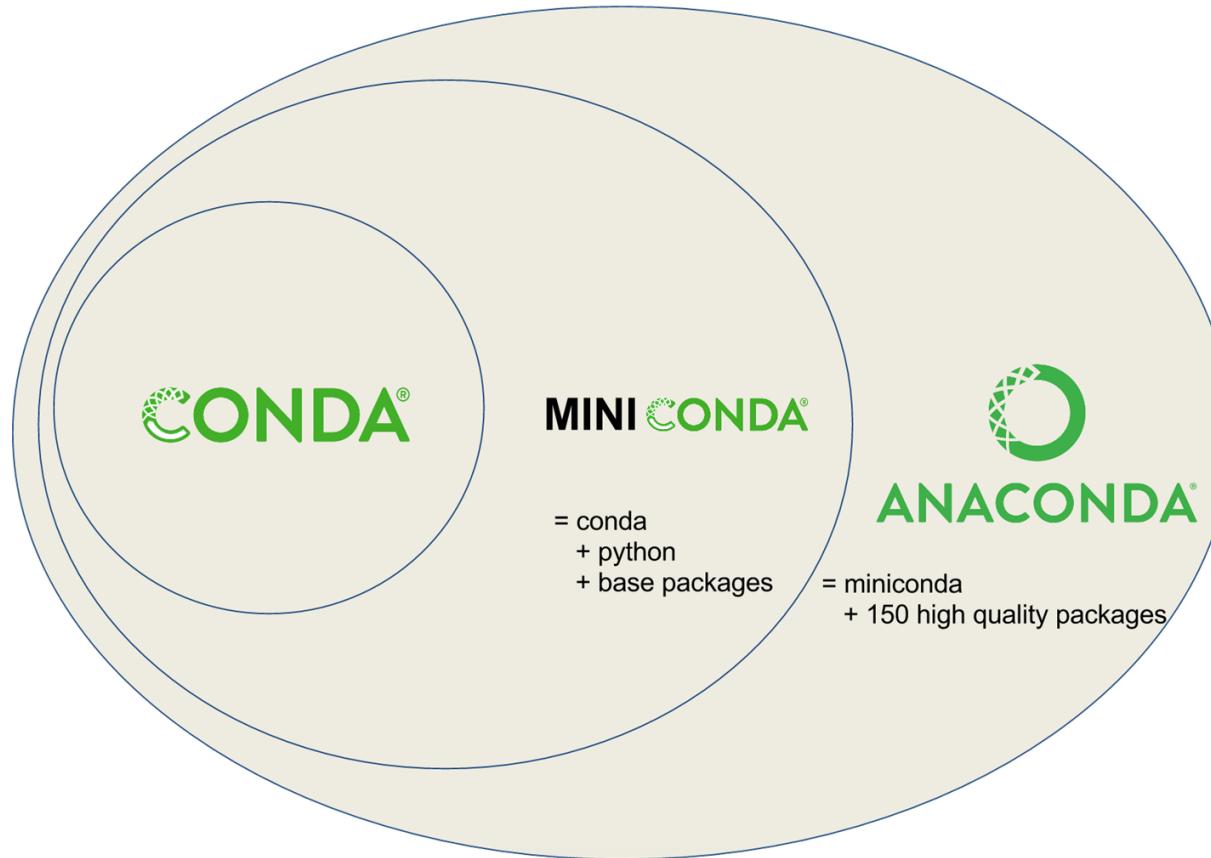
建立分析環境

關於社會學

數據分析

傑華老師的煩惱

總結



Conda: 套件管理平台
Miniconda
Anaconda

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

環境設定

Python 安裝

[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

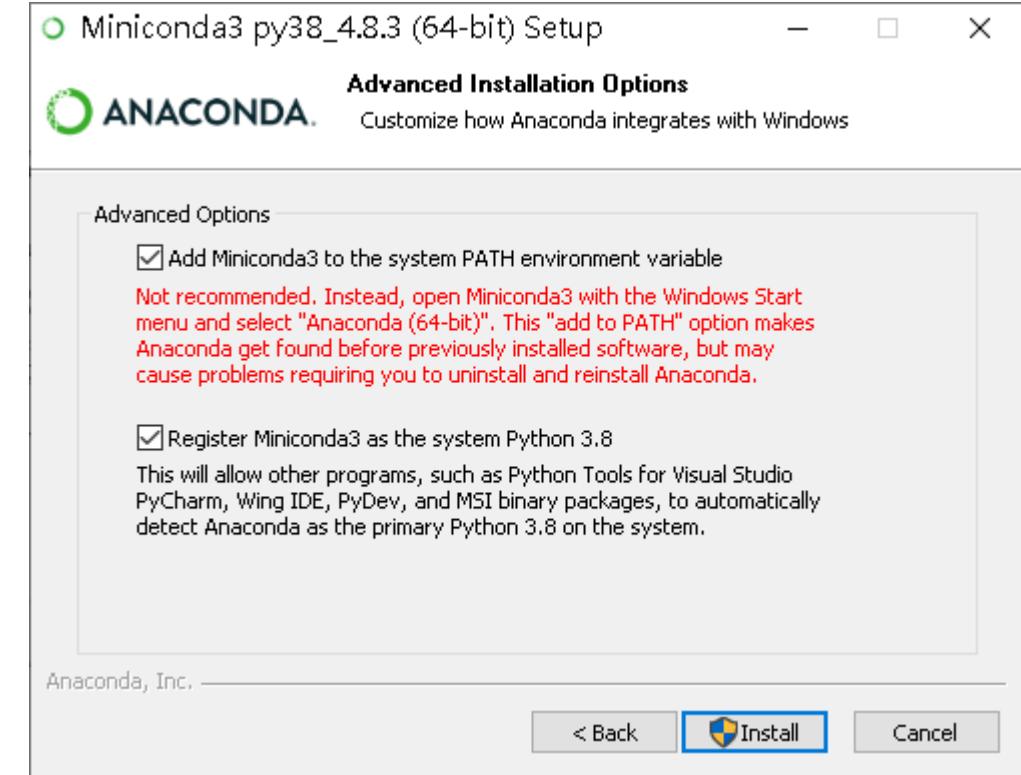
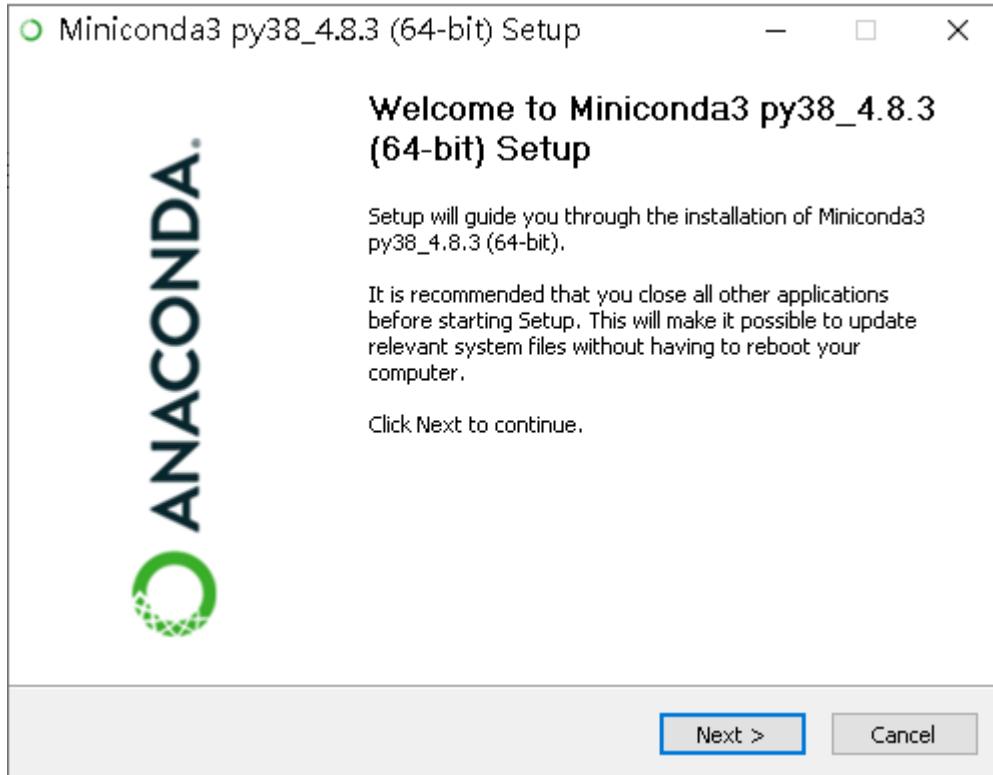
安裝Python

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做

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>"<
- 要喔，因為這可以幫你找到高薪的工作XD

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 ⑨

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 ⑨

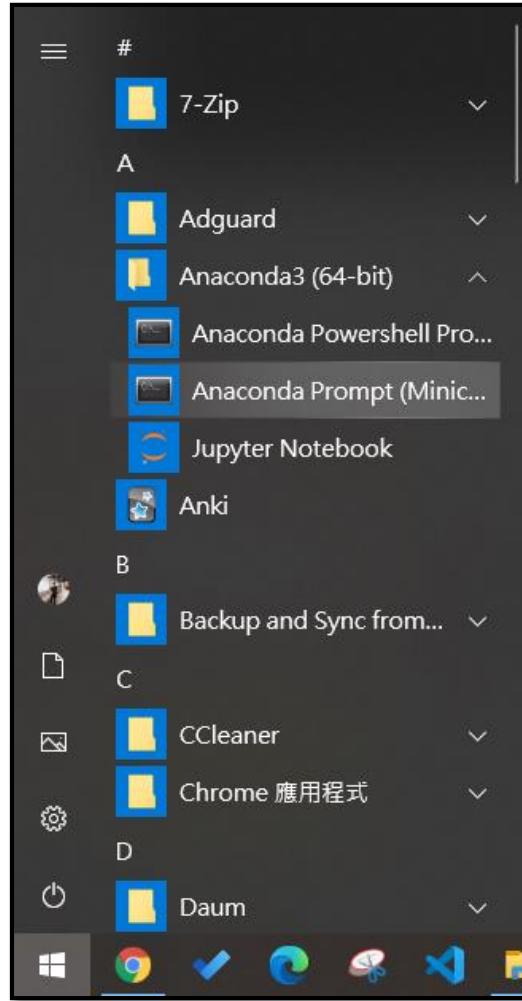
啟動jupyter

關於社會學

數據分析

傑華老師的煩惱

總結

A screenshot of an Anaconda Prompt window titled '系統管理員: Anaconda Prompt (Miniconda3)'. The window shows the command '(base) C:\Users\TL_Yu>jupyter notebook' entered and partially executed. The background of the window is black.

去年好像有開發一個爬蟲程式
但我忘記具體的流程跟方法了><

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', '家教', '外包', '獵才', '掌聲', '更多', '科技大廠聯合招募→搜尋薪優工作'. Below the bar, the main header reads '104 人力銀行' with sub-links for '學生', '新鮮人', '上班族', '主管職', '中高齡'. To the right are '求職登入/註冊' and '企業徵才' buttons. A large search bar at the top center contains the placeholder '關鍵字(例：工作職稱、公司名、技能專長...)'. Below it, a dropdown menu shows '地區' and '職務類別'. An orange '搜尋' button is to the right. The main content area features several job search filters: '金融業急徵', '無經驗可', '月薪4萬元起', '最新工作', '畢業生請進', '行政', '國外的業務', '百大', '英文', '福利好公司', '打工兼職', '外商專區', '中高齡', '科技園區', '上市上櫃', '學生實習', '日文', '薪水很滿意', '搶手工作', and '接受身心障礙'. To the right, there are links for '找公司', '公司評論', and '企業品牌大賞'. Below these filters, there are links for '職務找工作', '職涯診所', '履歷診療室', '職場萬象', '競爭力分析', '職務大百科', '地圖找工作', and '白皮書'. On the left, there's a featured article with a woman's photo and the title '服務業轉職怎麼做?' and a sub-section about '疫情衝擊' and '工作多年的我，該如何自救？專家幫你解惑'. On the right, there's another section titled '全民享補助 美日雙語輕鬆講!' with a '立即申請' button. The footer of the page includes a copyright notice: '科系、職務與待遇：大數據時代下的社會學與勞動市場' and an email address: 'Mail: TLYu0419@gmail.com'.

- 可以透過關鍵字、地區、職務類別搜尋資料
- 確認網站結構，檢查request訂位資料

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](#)

The screenshot shows the 104.com.tw job search interface. At the top, there's a navigation bar with links like '104人力銀行', 'My104', and '搜尋薪優工作'. Below the bar is a search form with fields for '關鍵字' (Keywords), '職務類別' (Job Type), and a large orange '搜尋' (Search) button. There are also checkboxes for '只搜尋職務名稱' (Search by job title only) and '切換高階職類' (Switch to advanced job type). Further down 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'. A '清空條件' (Clear filters) and '檢視或訂閱條件' (View or subscribe to conditions) button are also present. The main content area displays a job listing for 'CCPL BDO - 台北信貸業務【挑戰高獎金的工作在這裡!!】' from '渣打國際商業銀行股份有限公司'. The listing includes details like '台北市松山區 | 1年以上 | 專科', '月薪30,000~100,000元', and '員工3000人'. To the right of the listing is a sidebar with a '強力放送' (Powerful Broadcast) section containing a message about entrepreneurship and a '新版介紹' (New version introduction) button. Other buttons in the sidebar include '儲存' (Save) and '應徵' (Apply).

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...

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, 搜尋薪價工作, 帳號, 註冊, and 登入.

The main search form has '關鍵字(例：職稱、公司名、技能專長...)' and '台北市' selected as the location. 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 '儲存' (Save) and '應徵' (Apply).

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, requiring no experience or education. It lists responsibilities in English, office software proficiency, and presentation creation. The salary range is 月薪25,000~30,000元 and it has 20 employees. There are buttons for '儲存' and '應徵'.

The second listing is for ((雙北市))內湖區-展示中心(日班) at 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, requiring no experience or education. It lists responsibilities for managing visitors, vehicle entry control, and reporting site conditions. The salary is 月薪41,500元 and there are 1000 employees. There are buttons for '儲存' and '應徵'.

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

往下滑的時候可以看到上面網址的改變

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 Network tab in Google Chrome DevTools. The timeline at the top shows several requests being made over time. The main list of requests shows multiple entries for 'list?ro=0&area=6001001000&order=11&asc=0&...', indicating a scroll operation. The right-hand panel displays the details of one specific request:

- Request URL:** <https://www.104.com.tw/jobs/search/list?ro=0&area=6001001000&order=11&asc=0&page=24&mode=s&jobsource=2018indexpoc>
- Request Method:** GET
- Status Code:** 200 OK
- Remote Address:** 122.147.53.67:443
- Referrer Policy:** strict-origin-when-cross-origin

The Response Headers section lists various HTTP headers.

透過F12監聽網頁的互動方式

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 Network tab in Google DevTools for a job search page. The requests listed are all variations of the URL `list?ro=0&area=6001001000&order=11&asc=0&...`. The response pane displays a JSON object representing a successful API call. The JSON structure includes fields like `status`, `action`, `data`, `statusMsg`, and `errorMsg`. The `data` field contains a list of job items, each with details such as `jobType`, `jobNo`, `jobName`, `appearDate`, `applyCnt`, `applyDate`, `applyDesc`, `coIndustry`, `custName`, `custNameRaw`, `custNo`, `descSnippet`, `description`, `isApply`, `isSave`, `jobAddrNo`, `jobAddrNoDesc`, `jobAddress`, `jobName`, `jobNameRaw`, `jobNameSnippet`, `jobNo`, `jobRole`, `jobType`, `jobsource`, `landmark`, and `lat`.

找到資料的位置

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

人力銀行網路爬蟲實作

No module name是什麼東西

關於社會學

數據分析

傑華老師的煩惱

總結

- 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

其實有更多資訊藏在職缺網頁，
但因為時間關係有興趣的再私
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
- 10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 with a concerned expression. 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

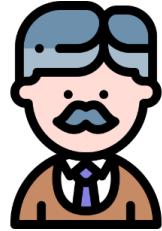
一周後 院長又找傑華老師吃飯



院長



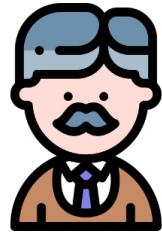
傑華



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20萬的職缺的資料呢！(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？



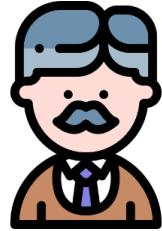
這是雇主在開職缺的時候，對於求職者的期望與勞動狀況，
可以看到職缺名稱、工作內容、要求技能與待遇等等資訊！



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的學生貢獻，
也會將分析成果轉給其他系所，
相信對他們也會是相當有價值的參考！

再回顧一下分析流程

資料清理

- 分析方式比較
- SPSS點選式，不用寫語法(適合一次性的分析，如學術研究)
- SPSS再處理資料很不方便
- Python/R 撰寫語法(企業需要定期執行程式)

- 社會學的議題很容易太大，要改變
- 對於學生的職涯規劃要有更積極的幫助

- 不是每個老師都喜歡計量研究，做報告前請先調查好老師的口味
- 學一下做簡報的技巧與設計的美感、配色
- 一定要學程式語言，分析的第一步是蒐集資料，
- 不要小瞧自己，但也不要委屈自己
- 要有獨立完成分析專案的能力，但你不能只會分析

- 發現問題，不要再慢慢研究了，要快快研究，採取行動，如果改變沒有達到成效，繼續想辦法改善！
- 鼓勵大家從小地方開始著手
- 透過怎麼樣的方式蒐集資料、做了什麼分析、帶來哪些改變
- 我覺得再小都是很有價值的事情！

- 所以社會學背景如果要走資料分析的領域要怎麼走?
- 建議往商業分析師的方向發展，技術能力也許沒這麼強，但一定要加強業務知識
- 如果有一些不服輸的精神，可以跟我一樣往資料科學家發展> “<https://blog.v123582.tw/2020/10/18/%E7%9C%9F%E3%83%BB%E8%B3%87%E6%96%99%E5%9C%98%E9%9A%8A%E8%88%87%E5%88%86%E5%B7%A5/?utm_source=Facebook_PicSee&fbclid=IwAR22NWGUjrv0MTzBOfsIGiddawZylnTWSFBRc_xhfsiVtlUFK7G_TyD8bJs

- Facebook社團
- Udemy

- 有想法給我，給系上
- 我會匿名、整理然後幫忙回饋給老師