

A man in a red shirt and tan pants is longboarding down a winding road through a mountainous landscape. The sun is low, creating a bright lens flare effect.

最後要確認每一頁右上有目前位置
字體

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
- 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

我的 社會學旅程

數據 應用

傑華老師的煩惱

總結

- | | | | |
|----------|-----------------|-------------|----------|
| • 關於社會學 | • 數據分析框架 | • 一份突如其來的工作 | • To 系上 |
| • 畢業之後的事 | • 數據分析在做
什麼? | • 資料蒐集 | • To 學生 |
| • 職涯規劃 | • 反面的應用案
例 | • 職缺分析 | • 一些學習資源 |

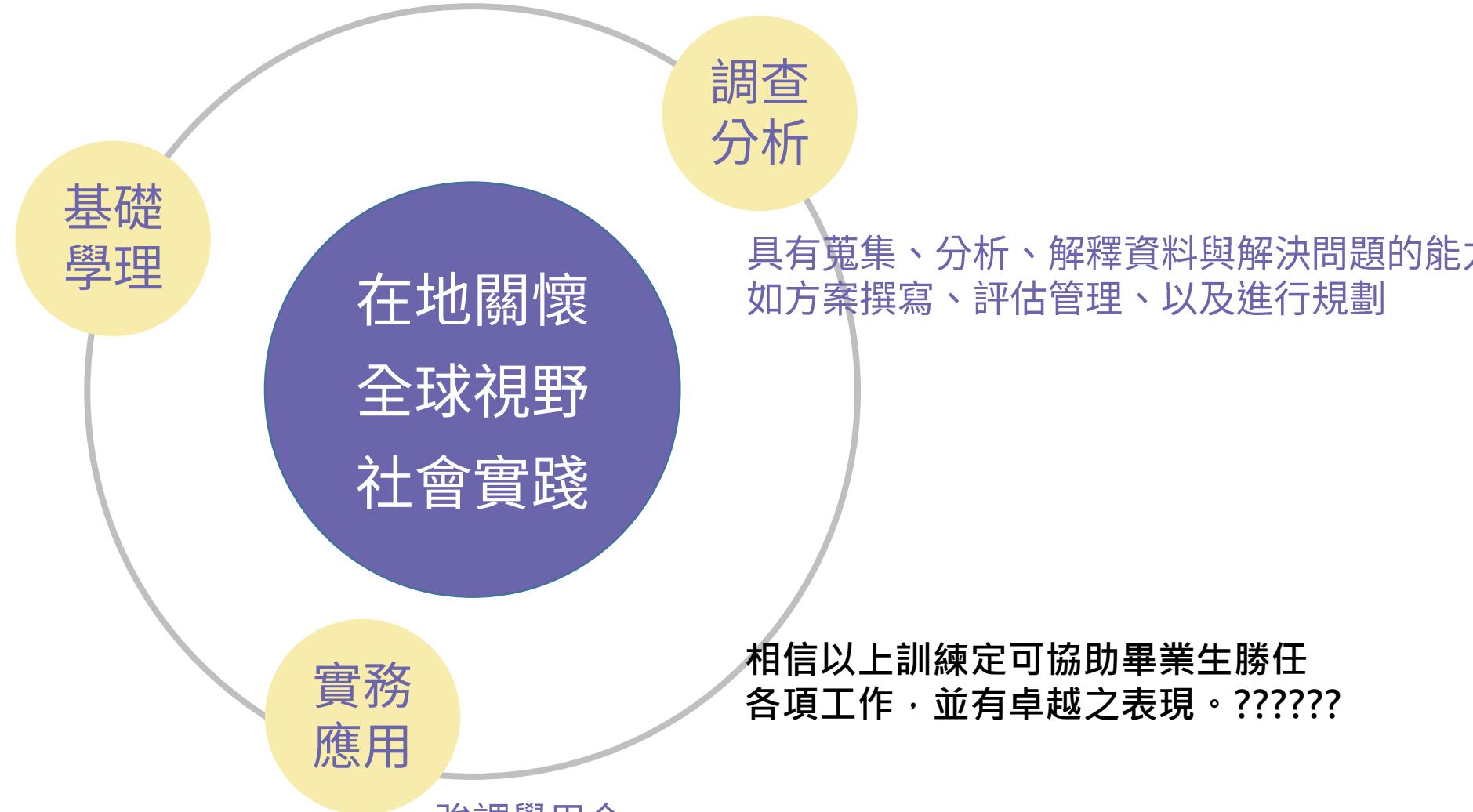
A man in a red hoodie and light-colored pants is longboarding down a paved road. He is leaning into a turn, looking back over his shoulder. The road curves through a landscape of green hills and mountains under a bright sky.

我的社會學旅程

- 關於社會學
- 那畢業之後呢？
- 職涯規劃

社會系課程規劃理念

強調基礎知識之掌握，
並培養思辨與批判的能力



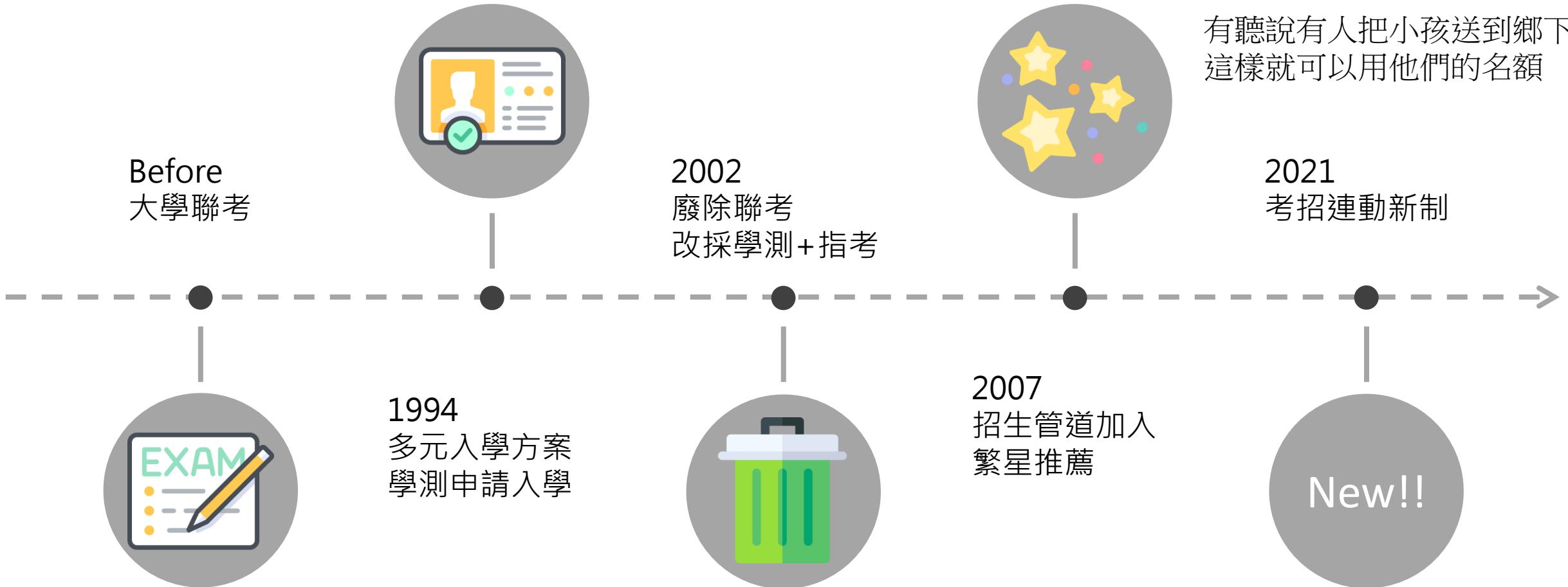
<https://spa.ndhu.edu.tw/p/412-1059-17584.php>

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

到底社會學在講什麼>”<

教育改革歷史

教改的目的是希望讓學生獲得的教育機會均等，但…有嗎？



- 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.

來分享一些社會分析的案例



了候選人找奇怪的人
691% E6% 9

/28/CHUNGYANGYEH/

臺灣年金制度改革之內涵與思考

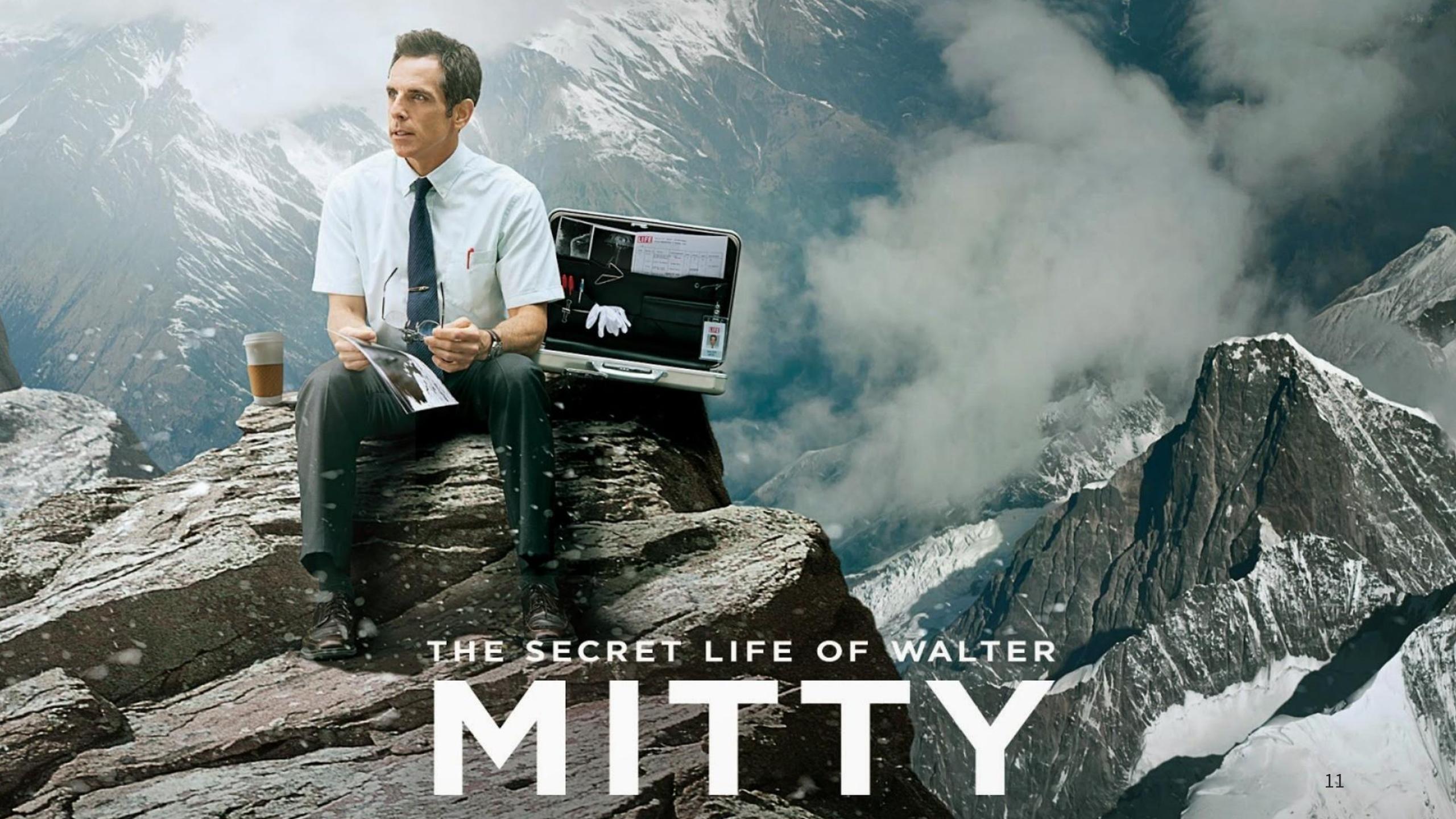
<https://twstreetcorner.org/2017/04/18/hongyonglin/>

來分享一些社會分析的案例



• 學運

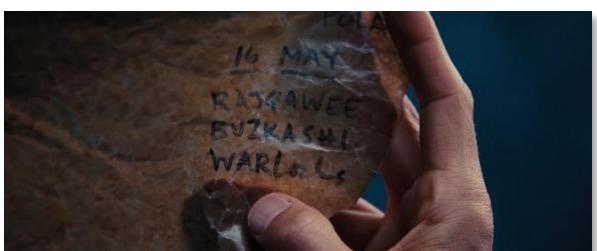
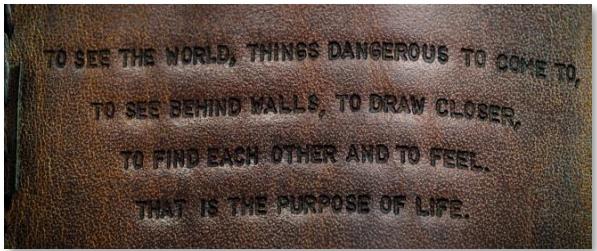
政治經濟結構
ORG/2014/05/09/WANGHONGZEN-4/



THE SECRET LIFE OF WALTER
MITTY

**那這些學了社會學的知識
對於職涯有什麼幫助呢?
難得的周末我們先來看個影片輕鬆一下XD**

所以為什麼要學社會學？



那麼職涯規劃呢？

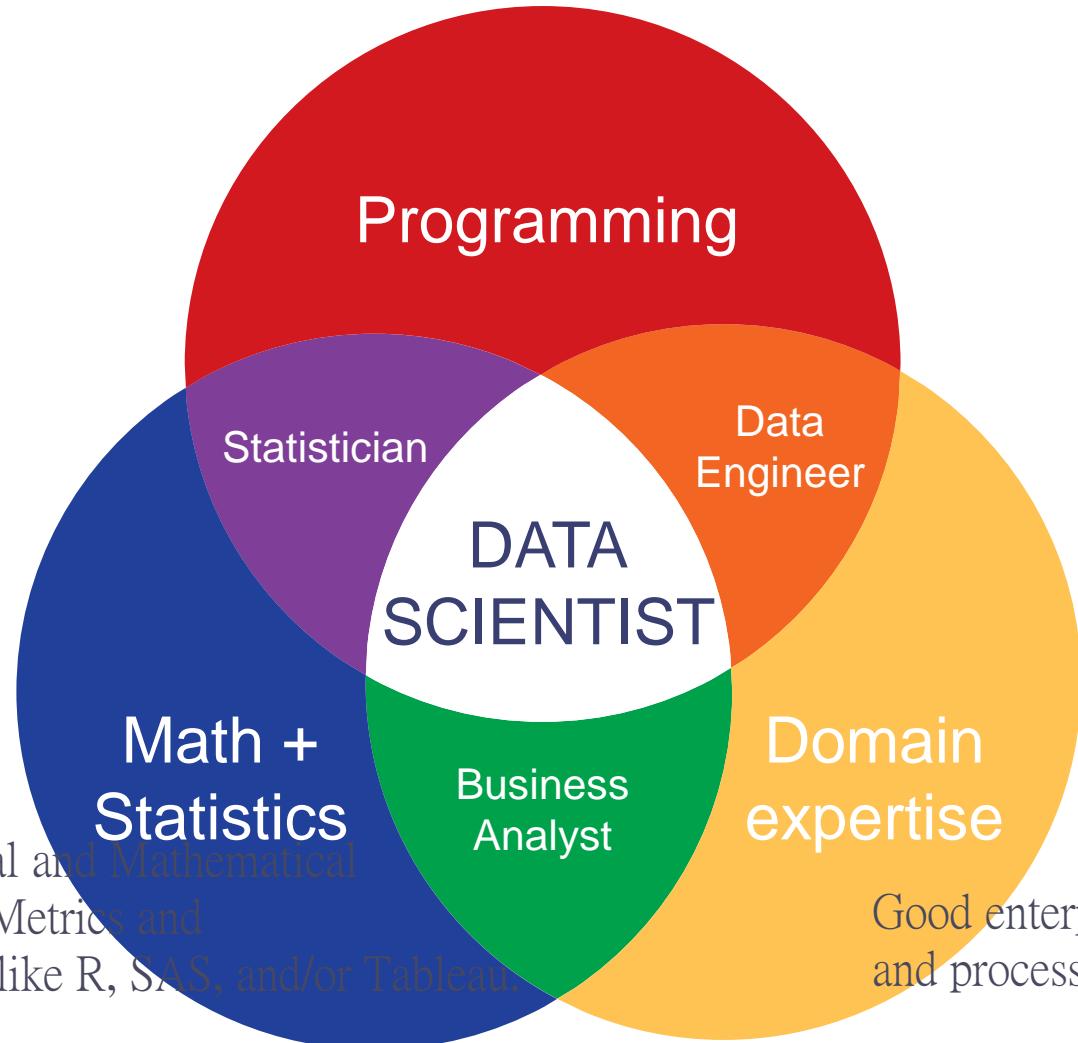
職涯規劃 看看畢業後的學長姐們



<https://www.facebook.com/photo?fbid=3133677146722093&set=gm.3046305698822041>

在資料科學的領域中有哪些角色

<https://www.zhihu.com/question/23946>



To project your data you need Statistical and Mathematical knowledge (Linear Algebra, Calculus, Metrics and Probability). Learn some popular tools like R, SAS, and/or Tableau.

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

- You need Computer Software Programming and coding skills to/for:
- *Process your data* by developing tools, Data Science models and Algorithms such as *Decision Trees*, *Random Forest*, *Support Vector Machines*, *Neural Networks*, and *R* or *Scala*.
 - *Design, Manage, and Backup Data*: RDBMS: Hadoop, MySQL, PostgreSQL, Oracle, and MongoDB; NoSQL: MongoDB, Redis, and Cassandra.
 - *Query large datasets*: SQL and NoSQL.
 - *Containerize*: Kubernetes and Docker.
 - *Version Control*: Git and GitHub.
 - *Cloud Server Architecture*: AWS, Azure, and Google Cloud Platform.
 - *Build ML Algorithms*: You need *Data science concepts* such as *Machine Learning*, *Statistical Inference*, *Bayesian Methods*, *Model Selection*, *Feature Engineering*, *Visualization*, and *Predictive modeling*.
 - ML techniques: *K-nearest neighbors*, *Regression*, *Classification*, *Clustering*, *Dimensionality reduction*, and *Ensemble methods*.
 - *Data Science tools and frameworks*: Python, Scikit, NLTK, NumPy, Pandas, TensorFlow, PyTorch, Keras, and Jupyter Notebook.
 - *Computer Science fundamentals*: Data Structures, Algorithms, Performance Analysis, and Computer Architecture (I/O and memory tuning).

- 開個表格呈現不同角色要求的技能差異，
- 同時強調不是說我會一個就好，而是都要會，但哪方面的能力更強

沒有正確的路，
而是要找到自己喜歡的路

找到喜歡的路的方法

- 要找到自己喜歡什麼有很多種方法
- 性向測驗
- 跟老師聊聊
- 多回想一些平常在做什麼事情的時候會忙到忘記時間？
- 來上學長今天的課程(笑)

我們來看看 QA 吧!

A man in a maroon hoodie and light-colored pants is longboarding down a winding asphalt road. He is leaning into the turn, with his front foot on the board and his back foot pushing off. The road is surrounded by lush green hills and mountains under a bright, slightly cloudy sky.

休息~

- 10分鐘後回來>” <

A man in a red shirt and grey pants is running on a paved road. He is leaning forward with his arms pumping. The road curves to the right. In the background, there are green hills and mountains under a bright sky.

數據應用案例

- 什麼是數據分析？
- 數據分析的應用案例
- 大數據時代下的社會學有什麼挑戰

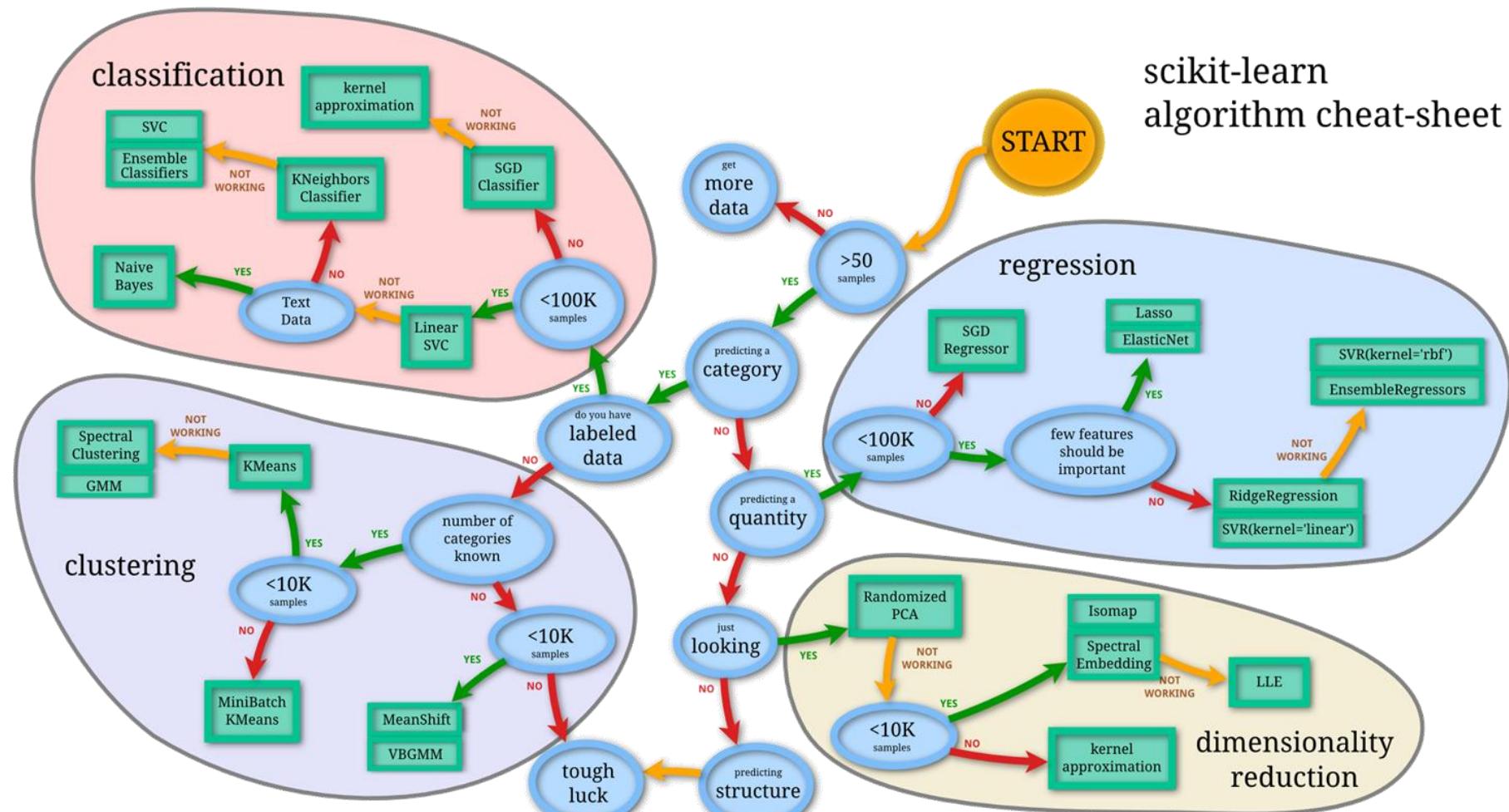
- 分析在做什麼？
- 敘述+預測
- 根據資料的尺度可以再進一步區分出分類、回歸、分群、降維

		有目標	無目標
	輸出類別	分類	分群
	輸出數值	回歸	降維

什麼是資料尺度

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...

如何選擇合適的演算法



分類模型案例

- (預測)分類：想知道100萬名客戶中哪些人會購買商品
- 為什麼要建模預測？單維度的標籤往往很難精準的預測
- 學期成績 過/不過 (老師的角度)

回歸模型案例

- (預測)回歸：想知道100萬名客戶中每位客戶各自會買多少錢
- 想知道客服中心接下來5天各自有多少通電話進線
- 為什麼要這樣預測？
- 學期成績會拿到多少分 (老師的角度)

降維模型案例

- (敘述)降維：想要萃取潛在結構???
- 通常用來做視覺化
- 系上的課程有很多，要精簡成績的構面(歷史、地理、公民)變成社會
- 物理化學變成自然

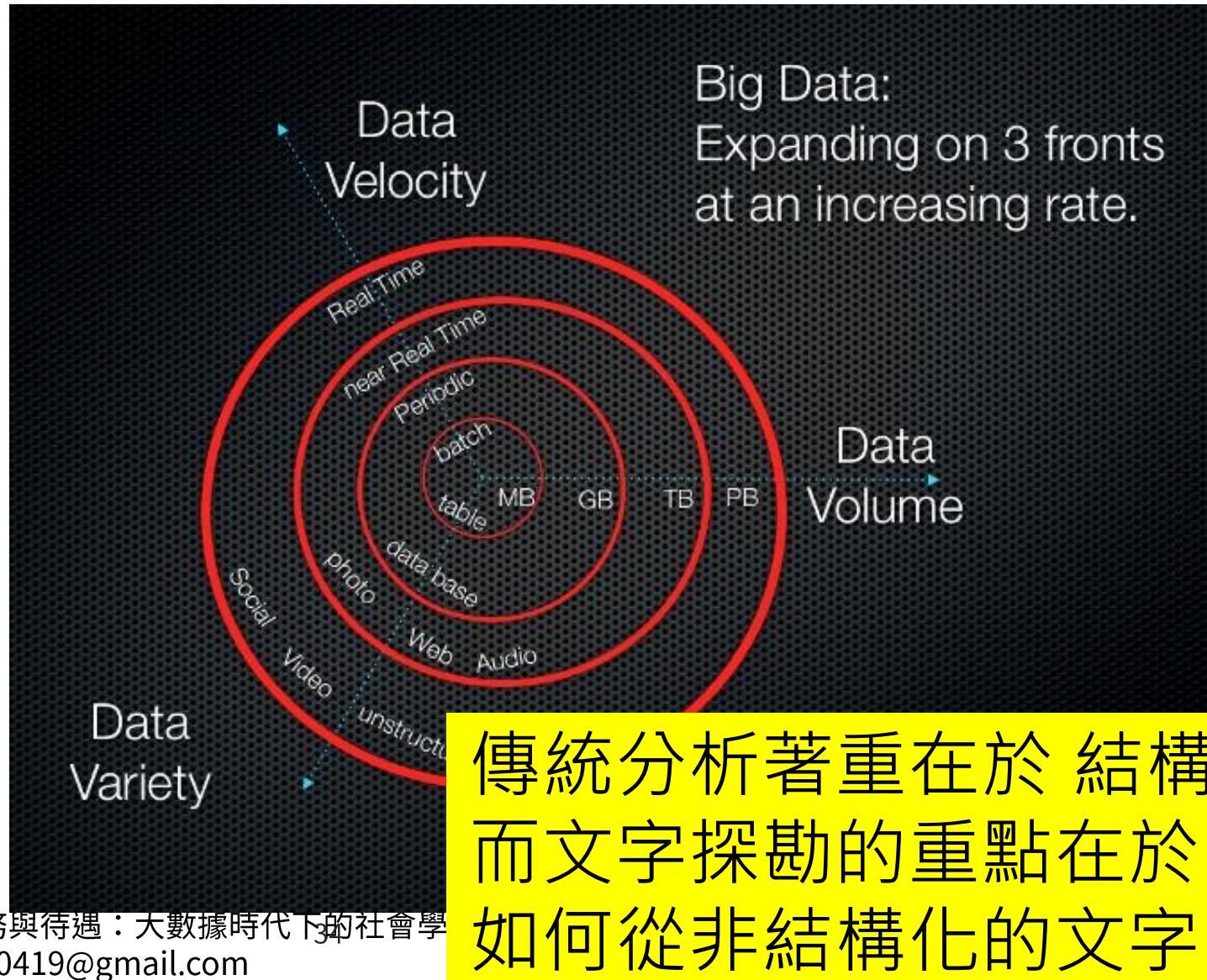
分群模型案例

- (敘述)分群：想知道客戶長什麼樣子???
- 班上學生的成績可以怎麼樣分群
- 我想要更直接一點，直接跟我說可以分成幾群，各群長什麼樣子

反面案例

- https://www.informationsecurity.com.tw/article/article_detail.aspx?tv=&aid=8311&pages=1

- 大數據的3個 V



✓ 資料量

✓ 形態

✓ 即時性

案例2

- Youtube推薦機制
- 狠準，但也造成了沉迷
- (教大家一招，如果男朋友沉迷在這上面可以偷偷幫她點不要推薦我這個頻道XD)

案例3

- 了解規則之後被玩弄規則
- 網軍、買讚

- 應用的反例：假新聞

- 買讚(還真的有人會買XD)

- 要強調不是說技術很可怕，所以我們要排斥他
- 反而我們也得要學會這些技術，雖然技術能力可能還是比不上資訊背景的人，但至少要具備基本的工程、分析能力!

總結為什麼變成反面

- 用的方式不對
- 沒有完美個模型，永遠都有誤差，但被錯誤預測的人是無辜的
- 預測會有準確度的問題(什麼是型一錯誤，型二錯誤)

什麼是型一型二錯誤

- 什麼叫做預測錯誤
 - 對於回歸模型叫做殘差
 - 分類模型有型1型2錯誤
-
- 颱風天，預測會不會停課
 - 預測成功，

到底大數據是好還是壞> “<

休息~

- 10分鐘後回來>” <
- 有問題可以在Slido上發問&幫別人點讚



我們來看看 QA 吧!

A man in a maroon hoodie and light-colored pants is longboarding down a winding asphalt road. He is leaning into the turn, looking back over his shoulder. The road is surrounded by lush green hills and mountains under a bright, slightly cloudy sky.

傑華老師的煩惱

- aaaa

就在上週
校長把傑華老師找進辦公室…

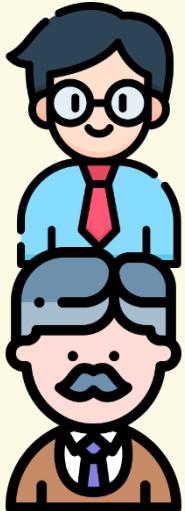


想要「討論」一下關於社會系學生的就業狀況(實際上是想要請傑華老師寫報告)，討論後希望傑華老師能幫忙了解以下幾個問題，並約傑華老師 1個月後繳交書面資料

市場上哪些產業/工作會想要找社會學系的學生？這些工作的工作內容是什麼？

不同科系職缺的就業狀況有哪些差異？社會學的優勢是什麼？

這幾年大數據很熱門，也有很多公司在做數位轉型，社會學系的學生在這個方面有哪些挑戰與機會？



...。(怎麼又把要給教育部的報告丟給我做，而且也不早點
跟我說，只給我 1個月怎麼來的及...
)



有問題嗎？

好的，沒問題

這邊用校長和老師的人頭來呈現對話

校長：傑華阿，你們社會系一直是以量化研究分析聞名的

傑華：這是系上師生一起努力的成果！(糟糕，該不會是有)

校長：你們去年是不是還有請一位畢業的校友回來分享大

傑華：對的，當時是請他回來分享網路爬蟲和文字探勘的

可以透過Python快速的爬取像是電商、社群、AP

校長：太好了，我這邊剛好有一份工作想要請你幫忙，

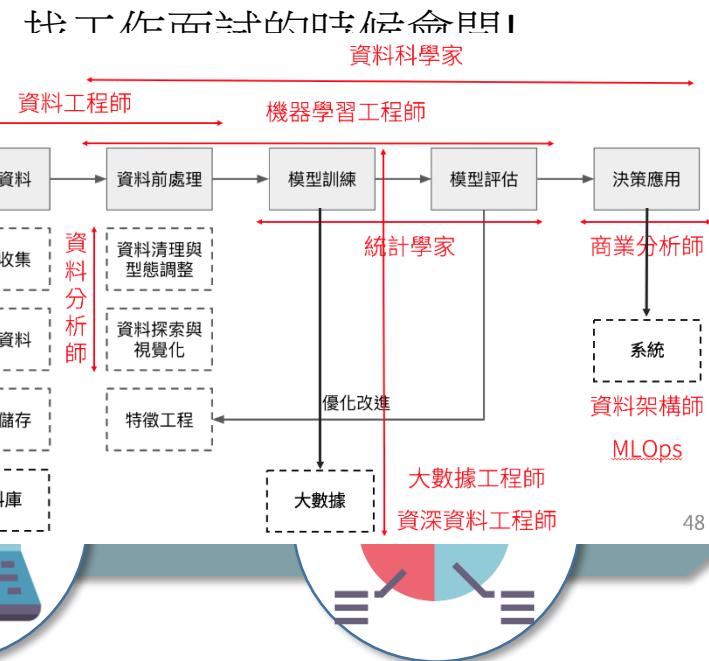
我這邊有接到教育部的報告，想請你幫我抓人力

傑華：...

在煩惱中傑華老師靈光一閃想
到了今天來參加課程的你們，
在課程中的你們學會了透過
Python 爬取資料與資料分析

分析流程介紹

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=IwAR22NWGUjrv0MTzBOfsIGiddawZyInTWSFBRc_xhfsiVtIUFK7G_TyD8bJs



目標

收集語料

- 社群網站
- 新聞文本
- 商品評論
- ...

預處理

- 同義詞
- 停用詞
- 特殊符號
- 斷詞
- ...

特徵工程

- 詞頻矩陣
- TF-IDF
- 詞向量
- ...

訓練模型

- 文本分類
- 情感分析
- 主題模型
- ...

確認目標

- 對社會有幫助
- 對企業有商機
- 很有趣(遊戲選哪個角色的勝率比較高)

知乎 首发于
王喆的机器

查看对话

宁肯 2019-05-05

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？
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

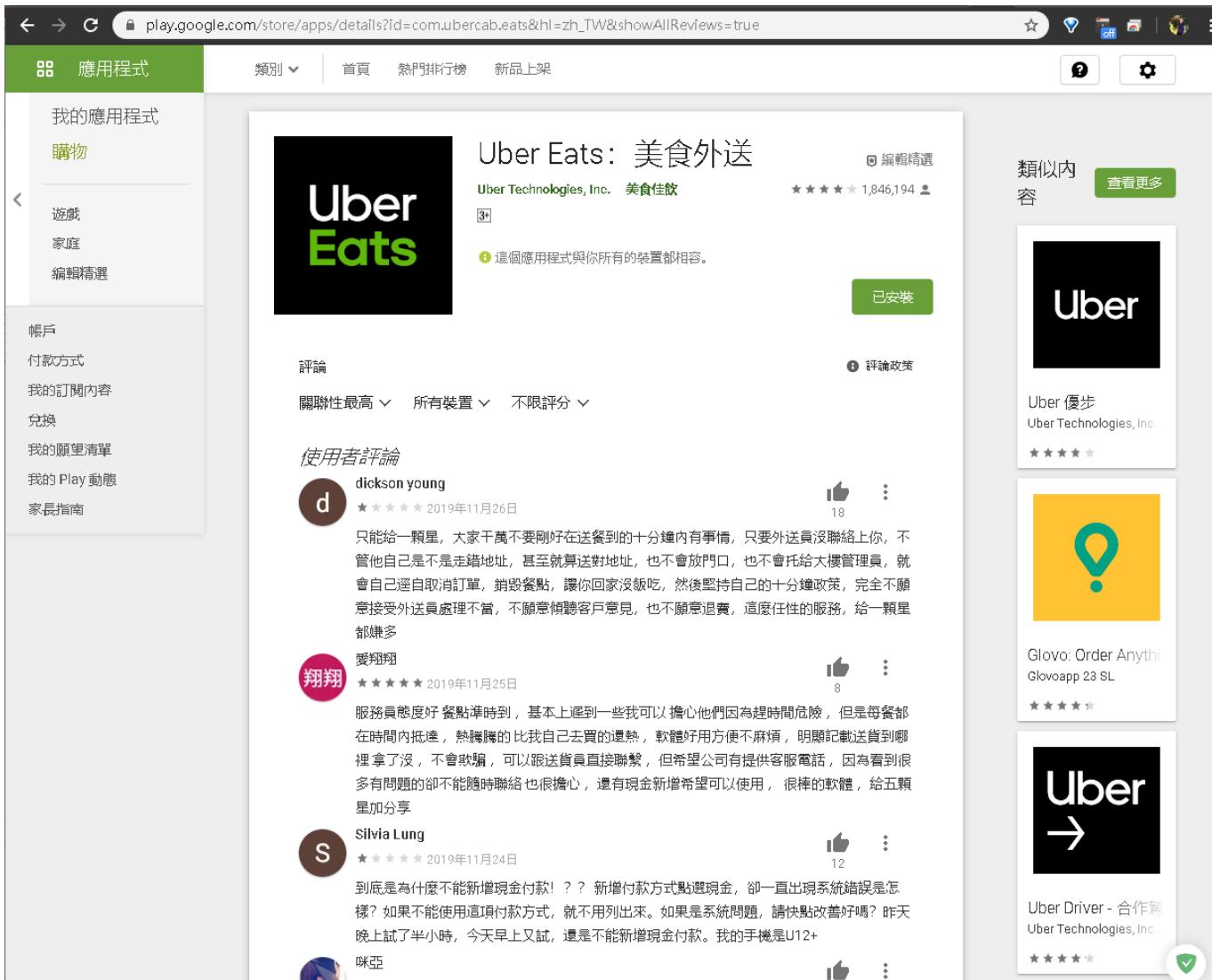
行者 回复 宁肯 2019-05-05

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）

資料蒐集 哪裡才有資料？

- 公司資料庫
- 政府公開資料
- 網路爬蟲

資料蒐集 我想優化公司的 APP



- Google Play 商店
- 資料內容
 - ✓ 名稱
 - ✓ 時間
 - ✓ 評價分數
 - ✓ 品論內容
 - ✓ 按讚數

資料蒐集 我想知道電商的價格與客戶的評論



- 購物網站
 - 資料內容
 - ✓ 商品名稱
 - ✓ 折扣數
 - ✓ 價格
 - ✓ 銷售量
 - ✓ 評分
 - ✓ 評論內容

資料蒐集 我想知道候選人的動態與網路聲量

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

[Facebook](#) 蔡英文 @iing

A screenshot of a Facebook post by Tsai Ing-wen. The post contains two paragraphs of text and two hashtags: #我的通關密語1203 and #鎖定直播. It has 17K likes, 916 comments, and 267 shares. Below the post are three comments from Tsai Ing-wen, each with a link to her Facebook event page. The comments are: "Like · Reply · 2h", "Like · Reply · 2h", and "Like · Reply · 2h". The last comment also includes a link to her website: <https://www.facebook.com/events/473640546595775/?ti=icl>. At the bottom of the post is a comment input field with placeholder text "Write a comment..." and several emoji icons.

- Facebook
- 資料內容
 - ✓ 發文人
 - ✓ 時間
 - ✓ 貼文心情數
 - ✓ 留言數
 - ✓ 分享數
 - ✓ 留言內容
 - ✓ ...

資料蒐集 我想知道勞動市場的職缺資訊

104.com.tw/jobs/search/?keyword=python&jobsource=2018indexpoc&ro=0&order=1

人力銀行 My104 家教 外包 找才 教育 更多 好企業哪裡找？快看企業品牌大賞

python 地區 職務類別 搜尋

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(3961) 全職(3716) 兼職(159) 高階(4) 其他 第 1 / 150 頁 符合度排序 強力放送

坐15年冷板凳！12強中美大職先發投手吳昇峰的故事

資料工程師 (Data Engineer)
和泰汽車股份有限公司
台北市中山區 | 2年以上 | 大學
1. 負責資料分析流程執行與優化，參與企業內部數據應用、顧客分析等實際專案執行。
2. 負責維護ETL系統與資料處理、運算等作業，能協助開發/維護資料分析介面。 *錄取後需至TOYOTA經銷商據點見學二個月
月薪 34,000~69,000元 上市上櫃 員工550人
儲存 電郵

AI架構師 / Algorithm Engineer / Data Engineer / Data Scientist
台泥企業團_臺泥資訊股份有限公司
台北市中山區 | 經歷不拘 | 碩士
<AI大數據工程師> 1.進行跨部門溝通，了解需求、分析議題並規劃Data Collection Schema 2.執行Data Exploration, Pre-stage visualization, Data Reduction, Feature Reduction與Machine Learning. 3.模組化與優化Data Collection /Cleaning Stage and 待遇面議 員工90人
儲存 電郵

11/21 Python Software Engineer
hyve solutions_海峰電腦股份有限公司 | 電腦系統整合服務業
桃園市龜山區 | 2年以上 | 大學
Hyve Solutions is looking for a talented junior/mid-level Python software engineer to help write Python programs. You are
待遇面議 外商公司 員工1003人
儲存 電郵

11/28 Python Developer
訊真科技股份有限公司 | 電腦軟體服務業
台北市大安區 | 2年以上 | 大學
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

• 人力銀行

• 資料內容

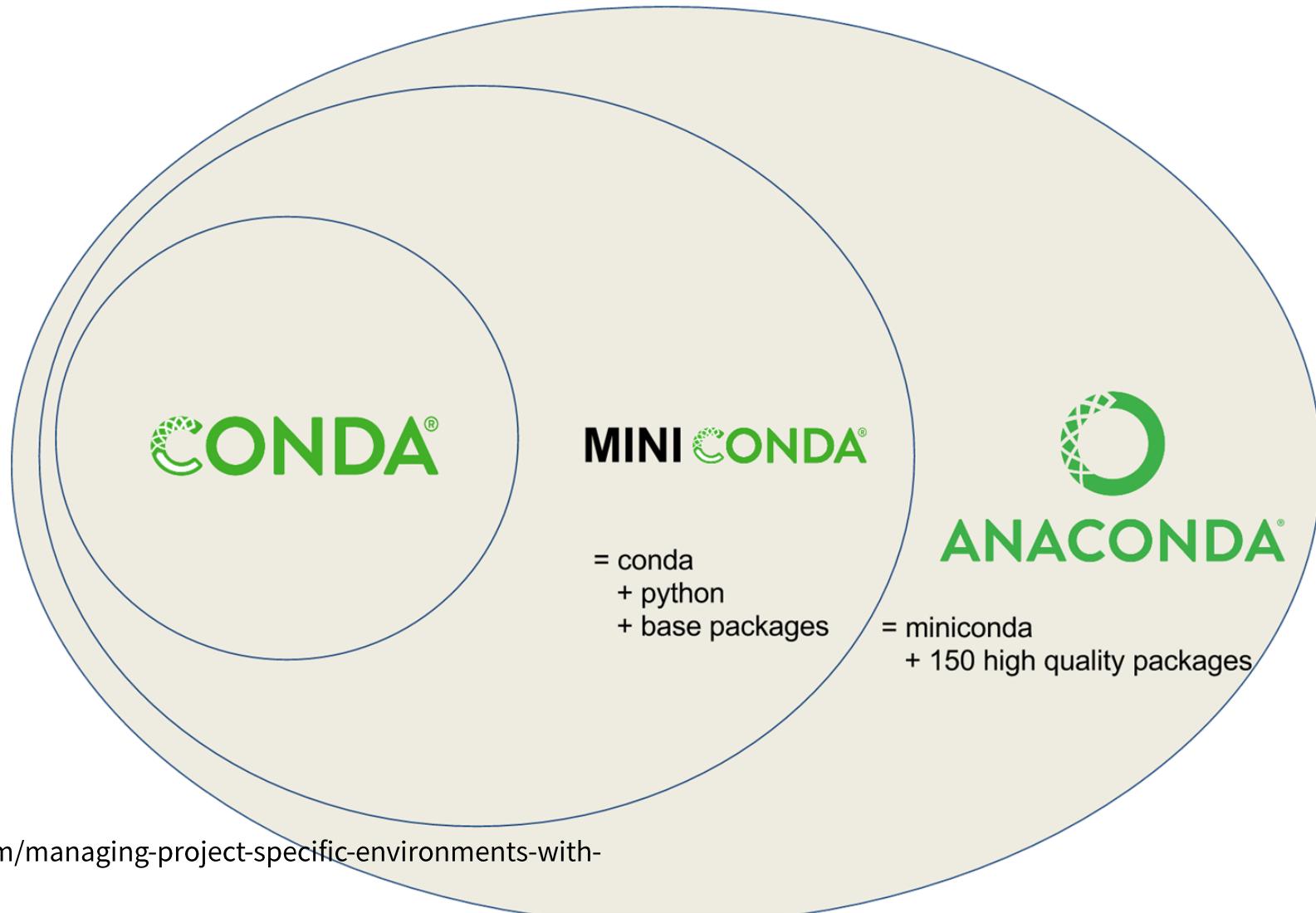
- ✓ 公司名稱
- ✓ 工作地區
- ✓ 所在地
- ✓ 職務內容
- ✓ 學歷要求
- ✓ 薪資
- ✓ ...

我不想做資料科學的工作 學這個對我會有幫助嗎？

- 會阿，你可以找到高薪的工作
(有人覺得這個不重要嗎？然後說不一定啦，因為每天3萬其實還滿有吸引力的XD)
- 而且找到學會這個還可以幫你找到高薪的工作
- 跟學弟妹互動，你覺得這個重要嗎?
 - 找到說重要的學弟妹，
 - 然後我反問妳社會系的，
你覺得賺錢跟對社會有幫助哪個重要？
 - 然後問為什麼？
 - 有些說重要

來實做一下網路爬蟲

建立分析環境

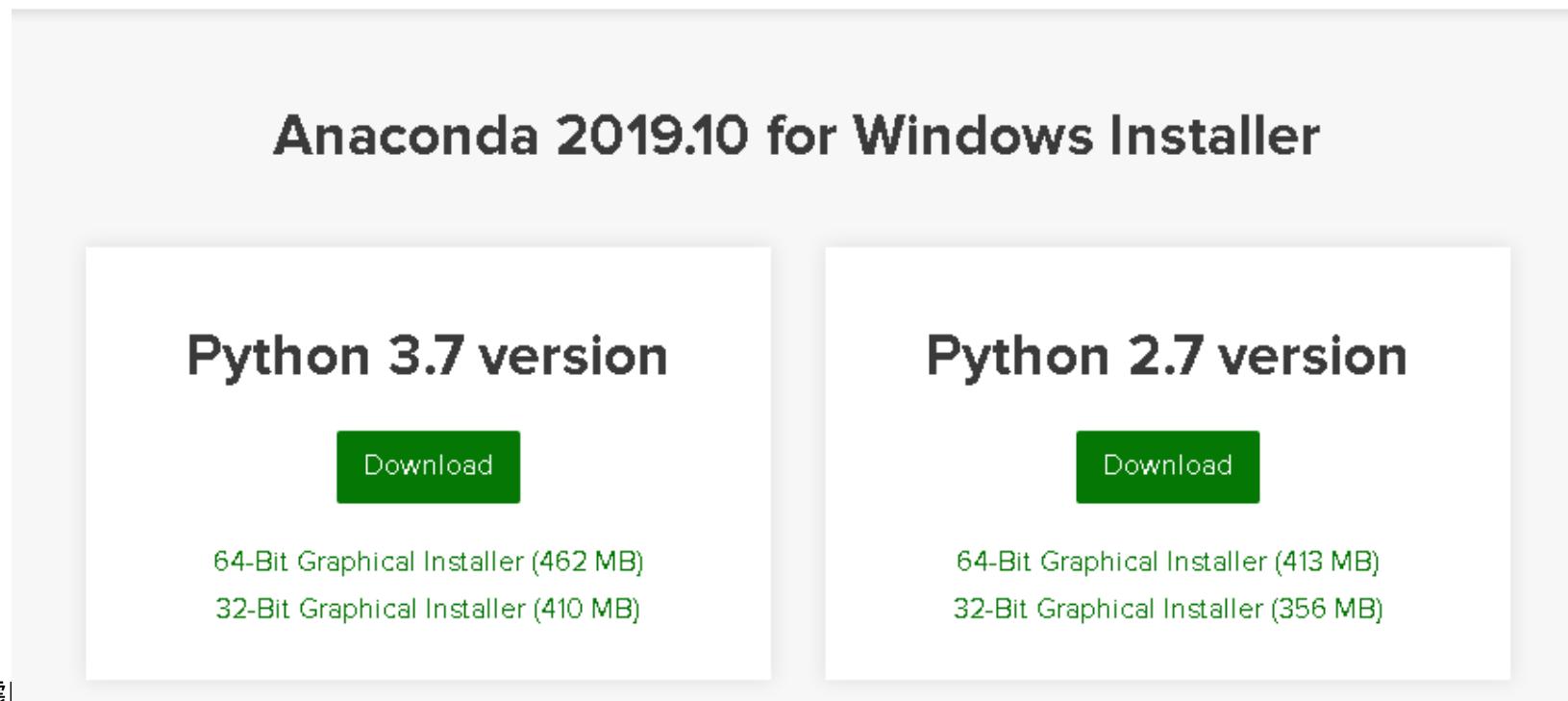


<https://towardsdatascience.com/managing-project-specific-environments-with-conda-b8b50aa8be0e>

環境設定

Python 安裝

- <https://www.anaconda.com/distribution/>



啟動jupyter

人力銀行網路爬蟲實作

帶他們看一下網站結構
練習爬外層的資料
反爬蟲機制介紹
加速

- 學長去年有來講怎麼爬蟲，但有點忘記該怎麼做了，我們來回想一下爬蟲要怎麼做...(時間回到1年前...)
- 可以講headers要怎麼設定，如果沒有放的話會怎樣
- 104有限制資料筆數的問題，所以要組合地區和職務來細分職缺
 - 地區代碼的表格是哪裡找到的
 - 職務代碼的表格怎麼找
 - json怎麼轉成DataFrame
- 怎麼寫發現規則 & 寫爬蟲程式
- [HumanResource_104_v2.ipynb](#)
- 學長碎碎唸：如果想要成為優秀的分析師，還是需要具備基礎的資料工程能力

安裝套件

- Conda install xxxx

費盡千辛萬苦，總算把職缺都
抓下來了！
我們可以看到抓下來的結果長
什麼樣子…

我們來看看 QA 吧!

A man in a maroon hoodie and light-colored pants is longboarding down a winding asphalt road. He is leaning into the turn, looking back over his shoulder. The road is surrounded by lush green hills and mountains under a bright, slightly cloudy sky.

傑華的煩惱II



傑華：報告校長，上次你交代的任務已經完成囉，我們花費了千辛萬苦，總算把職缺都透過大數據網路爬蟲的技術抓下來了(努力邀功一下)

校長：非常好!!既然你都抓好資料了，那我想要再多麻煩你一下，請你幫我作一下分析

傑華：OK

再回顧一下分析流程

資料清理

- 方法比較

	統計分析	大數據
目的	驗證假設	發現知識
方法	檢定因果關係	找出相關性
資料量	要多少樣本才有足夠的代表性	大量的資料與更大量的資料
成本	高	低
效率	低	高
類型	批次資料	即時資料

- 方法比較

	統計分析
目的	驗證假設
方法	檢定因果關係
資料量	要多少樣本才有足夠的代表性
成本	高
效率	低
類型	批次資料



• 資料結構

The screenshot shows the IBM SPSS Statistics Data Editor interface. A table is displayed with columns: id, zip, stratum2, qtype, year, year_m, b_dtime, b_dtm, and b_. The 'year' column is highlighted in yellow. The data consists of 735 rows, with the first few rows shown below:

	id	zip	stratum2	qtype	year	year_m	b_dtime	b_dtm	b_
1	100102	台北市中正...	都會核心	卷一	2007	96	7221030	7	
2	100106	台北市中正...	都會核心	卷一	2007	96	7161933	7	
3	100111	台北市中正...	都會核心	卷一	2007	96	8012100	8	
4	100112	台北市中正...	都會核心	卷一	2007	96	8011905	8	
5	100116	台北市中正...	都會核心	卷一	2007	96	8011617	8	
6	100120	台北市中正...	都會核心	卷一	2007	96	7161817	7	
7	100121	台北市中正...	都會核心	卷一	2007	96	7211627	7	
8	100125	台北市中正...	都會核心	卷一	2007	96	7242046	7	
9	100127	台北市中正...	都會核心	卷一	2007	96	7242135	7	
10	100131	台北市中正...	都會核心	卷一	2007	96	7222045	7	
11	100133	台北市中正...	都會核心	卷一	2007	96	7281027	7	
12	100136	台北市中正...	都會核心	卷一	2007	96	7242215	7	
13	100145	台北市中正...	都會核心	卷一	2007	96	7222145	7	
14	100151	台北市中正...	都會核心	卷一	2007	96	7182119	7	
15	100152	台北市中正...	都會核心	卷一	2007	96	7281135	7	
16	100156	台北市中正...	都會核心	卷一	2007	96	7221135	7	
17	100161	台北市中正...	都會核心	卷一	2007	96	8041725	8	
18	100162	台北市中正...	都會核心	卷一	2007	96	7170945	7	
19	100167	台北市中正...	都會核心	卷一	2007	96	8011807	8	
20	100302	台北市中正...	都會核心	卷一	2007	96	7231700	7	
21	100310	台北市中正...	都會核心	卷一	2007	96	7231001	7	
22	100312	台北市中正...	都會核心	卷一	2007	96	7310915	7	
23	100316	台北市中正...	都會核心	卷一	2007	96	7201945	7	
24	100317	台北市中正...	都會核心	卷一	2007	96	7261750	7	
25	100319	台北市中正...	都會核心	卷一	2007	96	7190915	7	

結構化資料

The screenshot shows a JSON file named 'records.json'. It contains several objects, each representing a record with a track ID, reporting date, longitude, and latitude. The data is color-coded for readability.

```

{
    "trackid": "AA-1234",
    "reported_dt": "12/31/2019 23:59",
    "longitude": "-111.12500000",
    "latitude": "33.37500000"
},
{
    "trackid": "BB-7890",
    "reported_dt": "12/31/2019 23:59",
    "longitude": "-113.67500000",
    "latitude": "35.87500000"
},
{
    "trackid": "CC-4545",
    "reported_dt": "12/31/2019 23:59",
    "longitude": "-115.57500000",
    "latitude": "37.67500000"
}

```

半結構化資料

The screenshot shows the DevTools of a browser (specifically Google Chrome) displaying the source code of a webpage. The page contains a large amount of unstructured data represented as JSON-like objects within the HTML structure. The data includes various fields such as 'trackid', 'reported_dt', 'longitude', and 'latitude', similar to the 'records.json' file.

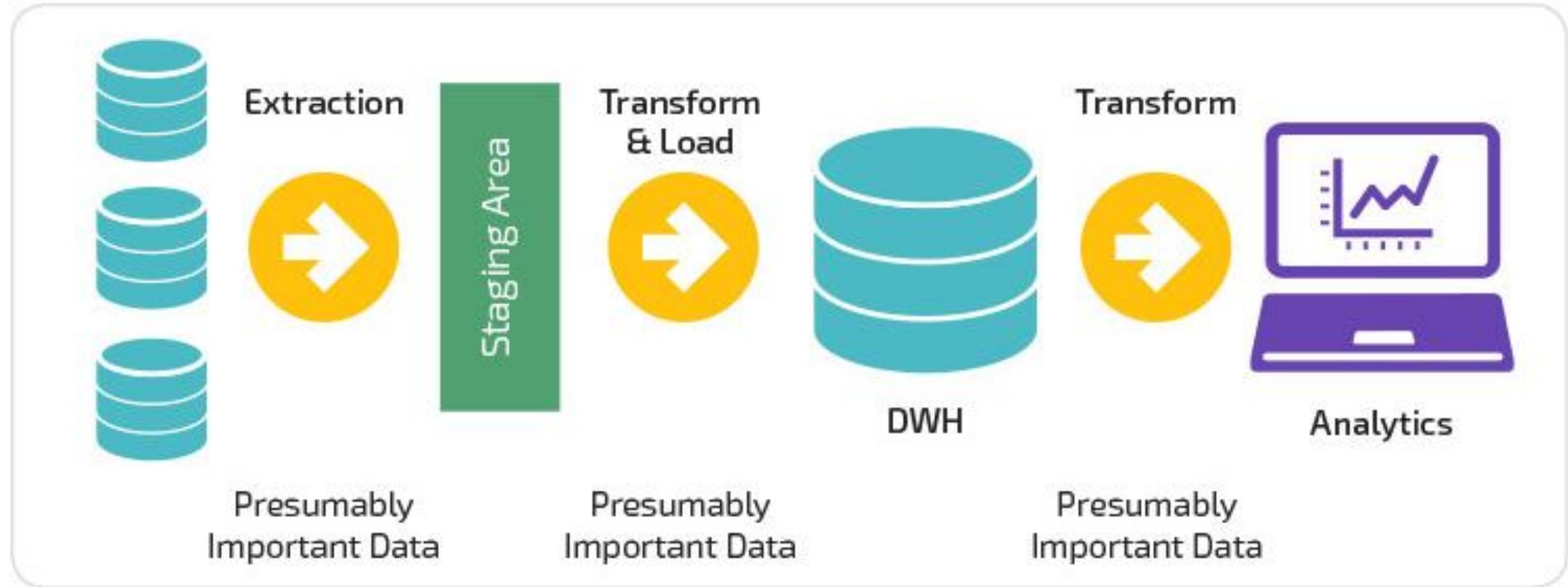
```

<body>
  <div id="root">
    <div class="LoadingBar"></div>
    <div></div>
    <main role="main" class="App-main">
      <div>
        <div class="SearchTabs SearchTab-bottomShadow" data-za-detail-view-path-module="SearchSwitchTabBar" data-za-extra-module="{"card":{"content":null}}"></div>
        <div class="SearchContainer">
          <div class="SearchMain" id="SearchMain">
            <div class="ListShortcut">
              <div class="List">
                <div class="data-za-detail-view-path-module="SearchResultList" data-za-extra-module="{"card":{"content":null}}">
                  <div class="Card SearchResult-Card" data-za-detail-view-path-index="0" data-za-extra-module="{"card":{"content": {"type": "Answer", "token": "91300972"}}, "attached_info_bytes": "Op0BCgtwbGfjZWhvbGRlchIgzt4ND1jMmQ0Y2yzNU1NDzMHzUxYnJ1Njc4YWZkNGEYBCIJHTE: HjUyKgk1Nz2zDcxTUyCDI0TgyNzgxOgkyOTEzNzY5HzdkF+eiseioiC1pkFn1bjm5og5beu! ABYAWABmAFAoAEBAqEA5AEBuAEAOAcH717CbdgBEuABBIAAbgCAA=="}"></div>
                  <div class="Card SearchResult-Card" data-za-detail-view-path-index="1" data-za-extra-module="{"card":{"content": {"type": "Answer", "token": "479307195"}}, "attached_info_bytes": "Op0BCgtwbGfjZWhvbGRlchIgzt4ND1jMmQ0Y2yzNU1NDzMHzUxYnJ1Njc4YWZkNGEYBCIJHTE: HjUyKgk1Nz2zDcxTUyCDI0TgyNzgxOgkyOTEzNzY5HzdkF+eiseioiC1pkFn1bjm5og5beu! ABYAWABmAFAoAEBAqEA5AEBuAEAOAcH717CbdgBEuABBIAAbgCAA=="}"></div>
                  <div class="Card SearchResult-Card" data-za-detail-view-path-index="2" data-za-extra-module="{"card":{"content": {"type": "Answer", "token": "576473399"}}, "attached_info_bytes": "Op0BCgtwbGfjZWhvbGRlchIgzt4ND1jMmQ0Y2yzNU1NDzMHzUxYnJ1Njc4YWZkNGEYBCIJHTE: HTE0Kgk1Nz2zDtz0TkycDNwNOTczNzxQgkzhDk1Q1NzZKF+eiseioiC1pkFn1bjm5og5beu! ABYAWABmAFAoAEBAqEA5AEBuAEAOAcH717CbdgBEuABBIAAbgCAA=="}"></div>
                  <div class="Card SearchResult-Card" data-za-detail-view-path-index="3" data-za-extra-module="{"card":{"content": {"type": "Answer", "token": "10A16601333"}}, "attached_info_bytes": "Op0BCgtwbGfjZWhvbGRlchIgzt4ND1jMmQ0Y2yzNU1NDzMHzUxYnJ1Njc4YWZkNGEYBCIJHTE: HTE0Kgk1Nz2zDtz0TkycDNwNOTczNzxQgkzhDk1Q1NzZKF+eiseioiC1pkFn1bjm5og5beu! ABYAWABmAFAoAEBAqEA5AEBuAEAOAcH717CbdgBEuABBIAAbgCAA=="}"></div>
                </div>
              </div>
            </div>
          </div>
        </div>
      </div>
    </main>
  </div>
</body>

```

非結構化資料

- ETL



網頁資料大多是非結構化的資料
在分析前需要先進行資料處理才能進行分析

- 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
- 不要小瞧自己，但也不要委屈自己

- 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

學習資源

- Facebook社團
- Udemy