

# **心理與神經資訊學**

## **(Psychoinformatics & Neuroinformatics)**

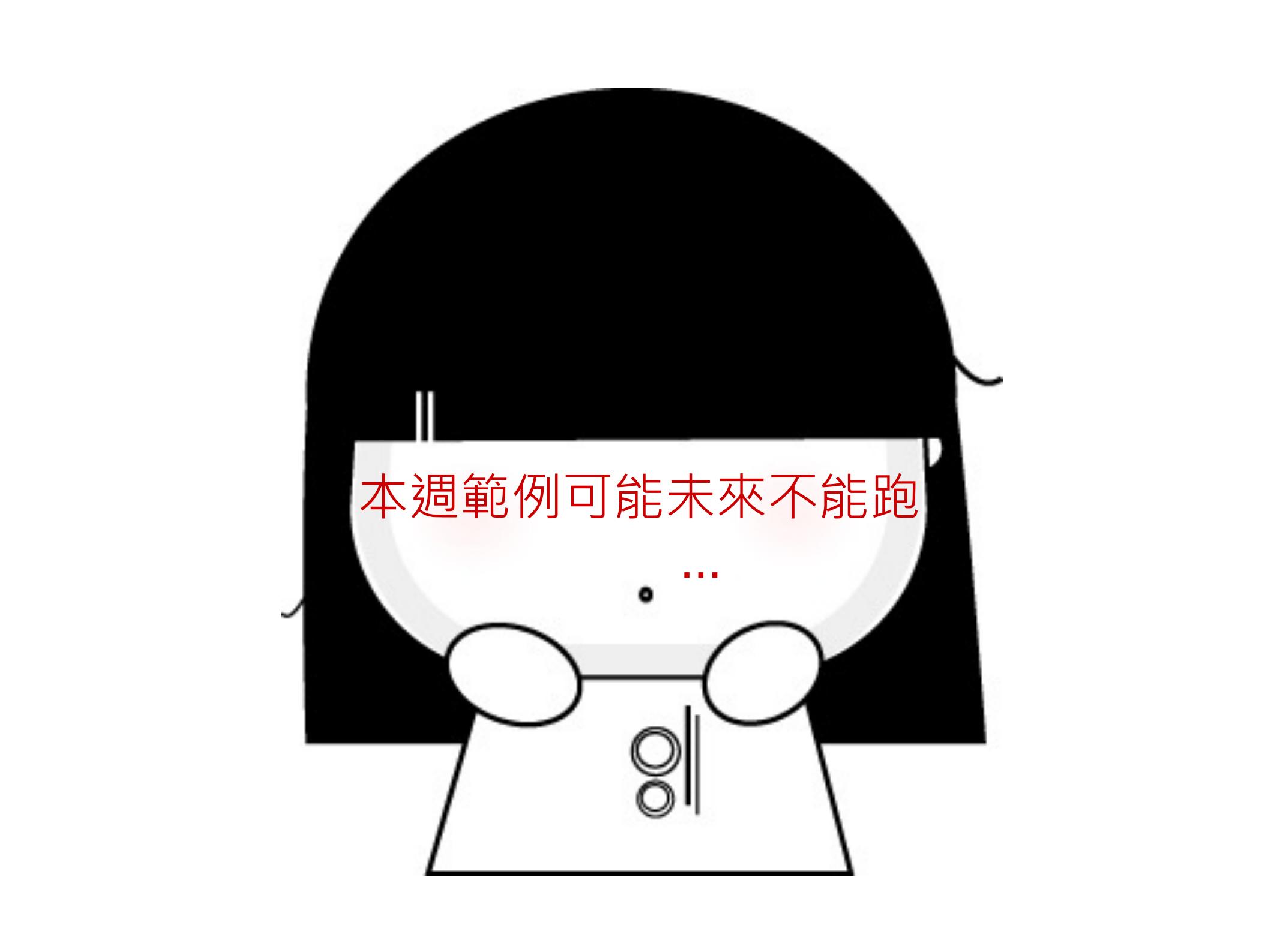
課號: Psy5261

識別碼: 227U9340

教室:彷彿在雲端

時間: 二789





本週範例可能未來不能跑

# 網路資料的搜集 (Web APIs)

# 今天需要安裝的模組

請在Jupyter Notebook下輸入：

**!pip install dcard-spider facebook-sdk tweepy**  
(以上分別是處理Dcard, FB, & Twitter需要的模組)

**!pip install oauth2 google-api-python-client**  
(以上分別是處理Dcard, FB, & Twitter的模組)

**!pip install --upgrade oauth2client**  
(以上是進一步更新處理Google服務的模組)



# What are API and SDK?

API=Application Programming Interface



黑盒子對外窗口

軟/硬  
體資源黑盒子

若把API看成函數，則類似Python Module的SDK (Software Development Kit) 則是一群APIs的集合

# 幻想文天堂Dcard

<https://www.dcard.tw/f/funny/p/226048661>

 搜尋有趣板、@人物  Q 註冊 登入 ▾

全部

▼ 分類看板

- 汽機車
- 語言
- 感情
- 女孩
- 美妝
- 穿搭
- 有趣

▶ 校園看板

## 套房隔音太差，常常聽到鄰居的...

3月25日 12:32

我租的套房雖然是水泥隔間  
但因為屋頂是輕鋼架所以隔音還是不太好  
沒聽過隔壁貓叫春  
只有一個不知道是隔壁還是樓下的女人很愛半夜唱歌

某天被她吵得不行  
一怒之下把wifi熱點的名字改成

別他媽半夜唱歌啊！

• 逐條顯示 4G 03:35 100%  

 29284  228  

# 神秘的Dcard APIs

<https://www.dcard.tw/service/api/v2/posts/226048661>

看板資訊	/v2/forums
文章資訊	/v2/forums/看板名稱/posts (照時間排序: popular=false) (照發文前後: before/after=文章編號)
文章內文	/v2/posts/文章編號 (照時間排序: popular=false)
文章內連結	/v2/posts/文章編號/links
文章內留言	/v2/posts/文章編號/comments (照留言前後: before/after=樓層編號)

# 用Python呼叫Dcard APIs



Web APIs大部分回傳的是json格式

```
import urllib.request,json
u='https://www.dcard.tw/_api/posts/226048661'
r=urllib.request.Request(u,headers={'User-Agent':''})
data=urllib.request.urlopen(r).read()
j_data=json.loads(data.decode('utf-8'))
print(j_data['media'][0]['url'])
for key in j_data.keys():
    print(key,':',j_data[key])
```

**Dict → json.dumps → String**  
**Dict ← json.loads ← String**

# Dcard有非官方的Python模組

更多使用細節請參考[此文](#)

```
from dcard import Dcard
def filter(metas):
    return [m for m in metas if m['likeCount'] >= 100]
#return [m for m in metas if '電影' in m['topics']]

d=Dcard()
f=d.forums('photograph')
m=f.get_metas(num='all',filter=filter) #list
p=d.posts(m).get(content=True)
r=p.parse_resource()
done,fails=p.download(r)
print('Got %d pic' % done if len(fails)==0 else
```

# 臉書上的按讚行為可預測個資 (1/3)

因為不同性別、年紀、族群的人喜好不同

1

## Users' Facebook Likes

55,814 Likes

58,466 Users

	art	cnn.com	(...)	BMW
User 1	1	1	...	0
User 2	0	1	...	1
User 3	1	0	...	0
(...)	...	...	...	...
User n	1	1	...	0

User – Like Matrix  
(10M User-Like pairs)

2

## Singular Value Decomposition

100 Components

58,466 Users

	Comp <sub>1</sub>	Comp <sub>2</sub>	(...)	Comp <sub>100</sub>
User 1	1.5	.7	...	-.9
User 2	.3	-.4	...	-.2
User 3	-.6	.1	...	4.7
(...)	...	...	...	...
User n	1.2	1	...	-.6

User – Components Matrix

3

## Prediction Model

Using Logistic or Linear Regression  
(with 10-fold cross validation)

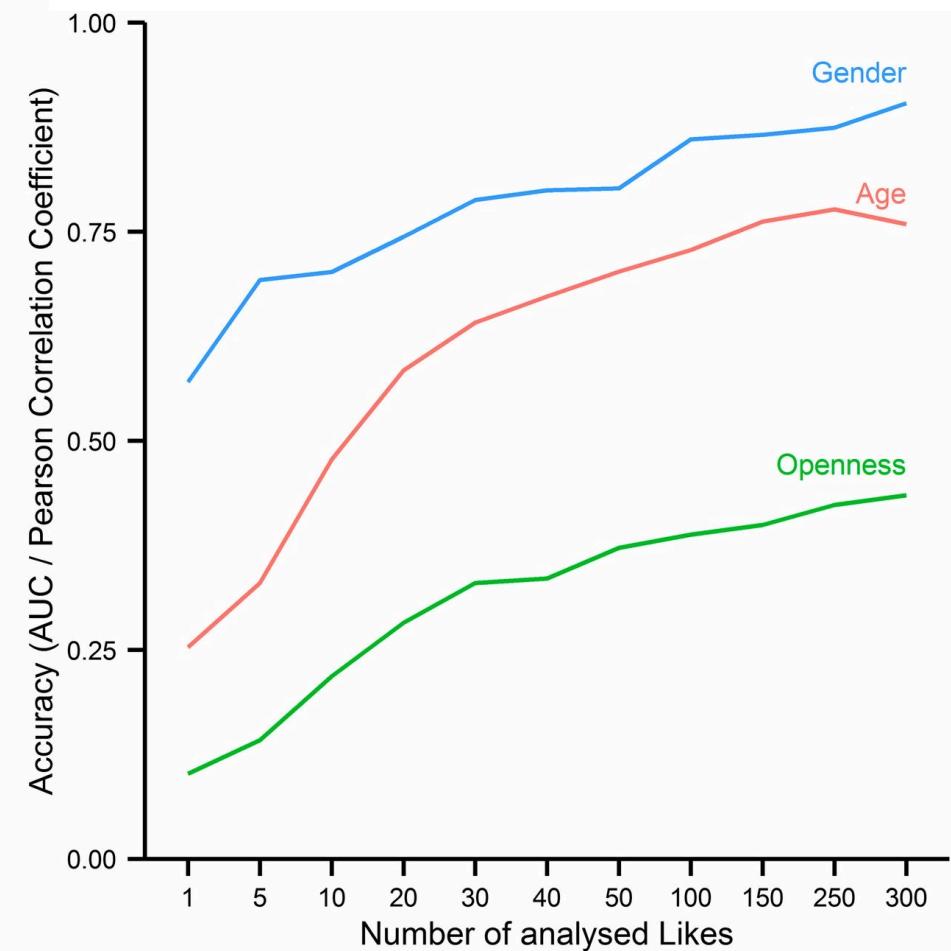
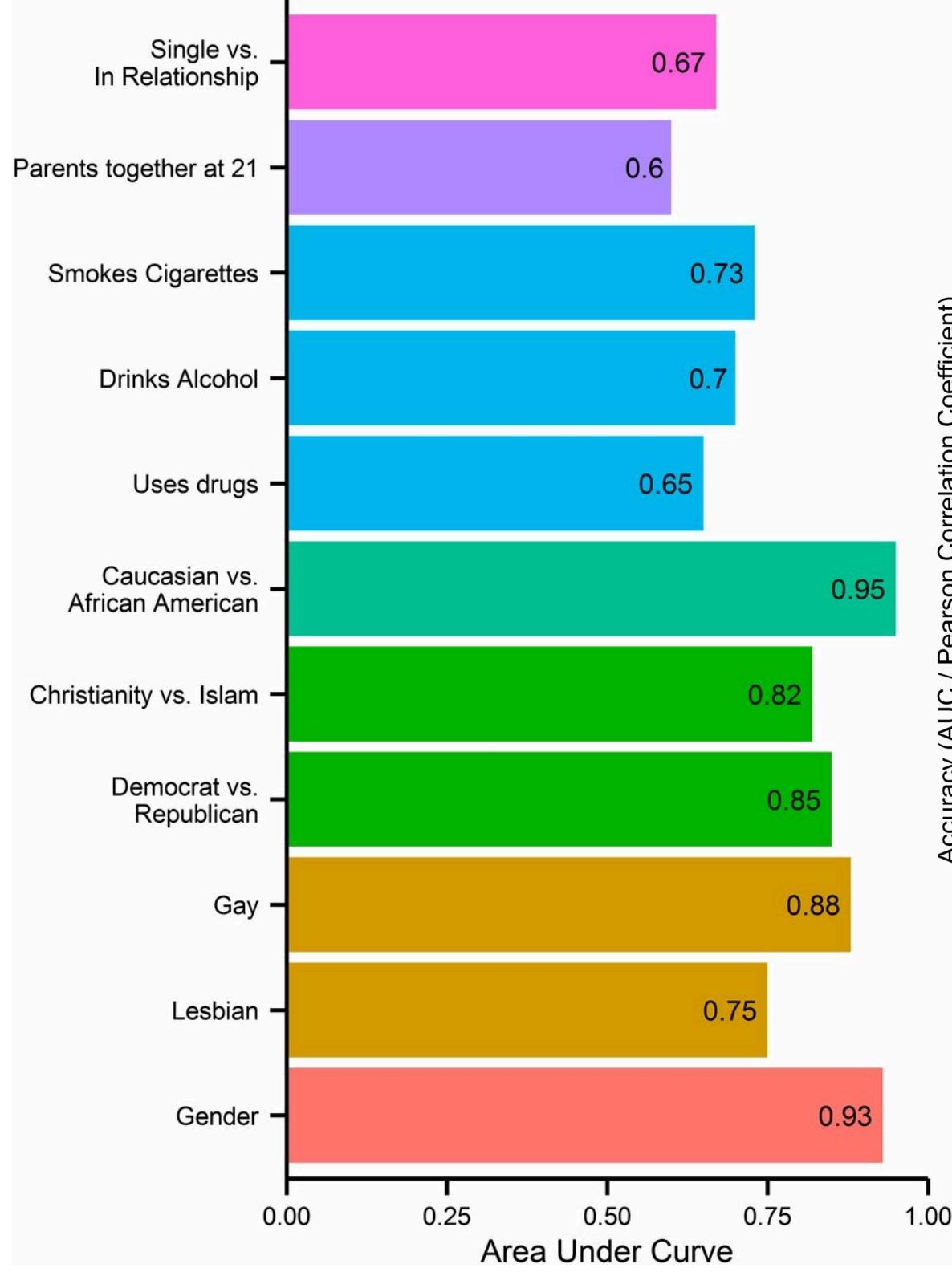
$$\text{e.g. } \text{age} = \alpha + \beta_1 C_1 + \dots + \beta_{100} C_{100}$$

### Predicted variables

Facebook profile: age, gender, political and religious views, relationship status, proxy for sexual orientation, social network size and density

Profile picture: ethnicity

Survey / test results: BIG5 Personality, intelligence, satisfaction with life, substance use, parents together?



# 臉書上的按讚行為可預測個資 (2/3)

且比人類的預測還要準

## PARTICIPANTS' PERSONALITY

Measured using 100-item IPIP Five-Factor Model questionnaire (for 70,520 participants)

	User 1	User 2	User 3	(...)	User X
Openness	2.1	1.9	...	...	...
Conscientiousness	4.2	2.1	...	...	...
Extraversion	1.9	3.2	...	...	...
Agreeableness	5.0	4.2	...	...	...
Neuroticism	3.2	4.3	...	...	...



90% of participants

1

Take personality scores and Likes of 90% of the participants and build linear regression models for the five personality traits using LASSO variable selection

## PARTICIPANTS' LIKES

Obtained from Facebook profiles

	User 1	User 2	User 3	(...)	User X
Running	1	0	1	...	...
Ford Explorer	1	1	0	...	...
Barack Obama	0	1	1	...	...
(...)	...	...	...	...	...



90% of participants

10%

Running  
Ford Explorer  
Barack Obama  
(...)  
Dancing

## LINEAR REGRESSION MODELS

A regression formula with a coefficient for each Like is generated for each of the five personality traits  
e.g. Openness =  $\alpha + \beta_1 * \text{running} + \beta_2 * \text{Obama} + \dots + \epsilon$



Openness	.3	.2	0	- .2
Conscientiousness	.7	.1	.6	- .7
Extraversion	.1	0	.1	- .2
Agreeableness	...	...	...	...
Neuroticism	...	...	...	...

2

Take the Likes of the remaining 10% of the participants and use the linear regression models to predict scores for the five personality traits

## COMPUTERS' JUDGMENTS

Made using participants' Likes

	User 1	User 2	User 3	(...)	User X
Openness	2.1	1.9	...	...	1.9
Conscientiousness	4.2	2.1	...	...	2.1
Extraversion	1.9	3.2	...	...	3.2
Agreeableness	5.0	4.2	...	...	4.2
Neuroticism	3.2	4.3	...	...	4.3



90% of participants

10%

Repeat 10 times to make judgments for all participants



Humans' Judgments



Self-ratings



Computers' Judgments

# 臉書上的按讚行為可預測個資 (3/3)

可以自己測試看看Apply Magic Sauce

The screenshot shows the homepage of the [Apply Magic Sauce](http://applymagsauce.com) website. The top navigation bar includes links for You, About Us, Research, Business, Documentation, and Contact Us. Below the navigation are three icons: two people, gears, and a building.

The main content area features a large button labeled "Apply Magic Sauce". To its right is a web browser interface showing the URL <http://applymagsauce.com>. Below this are two sections: "Personalise the Internet" and "A personalisation engine that accurately predicts psychological traits from digital footprints of human behaviour". A call-to-action button "Predict My Profile" is highlighted in yellow, along with a link "Don't want to use Facebook? Take the full test".

On the right side, there are two data visualizations: a bar chart titled "Education" and a pie chart titled "Relationship Status".

**Education** (Bar Chart):

Field	Percentage
Art	19%
Psychology	11%
Journalism	9%
Business	8%
Engineering	8%
Biology	8%
IT	7%
History	7%
Nursing	7%
Law	6%
Education	5%

**Relationship Status** (Pie Chart):

Status	Percentage
In a relationship	30%
Married	~25%
Single	~45%



# Facebook APIs

Facebook APIs的詳細介紹可以看[這裡](#)

原頁面：<http://www.facebook.com/taiwanpsy>

純資料：<http://graph.facebook.com/taiwanpsy>

[Explorer](#)可幫助我們找出還有何資料可看：

圖連接：

[https://graph.facebook.com/taiwanpsy?fields=albums.fields\(link\)&access\\_token=...](https://graph.facebook.com/taiwanpsy?fields=albums.fields(link)&access_token=...)

文敘述：

[https://graph.facebook.com/taiwanpsy?fields=posts.limit\(5\).fields\(description\)&access\\_token=...](https://graph.facebook.com/taiwanpsy?fields=posts.limit(5).fields(description)&access_token=...)



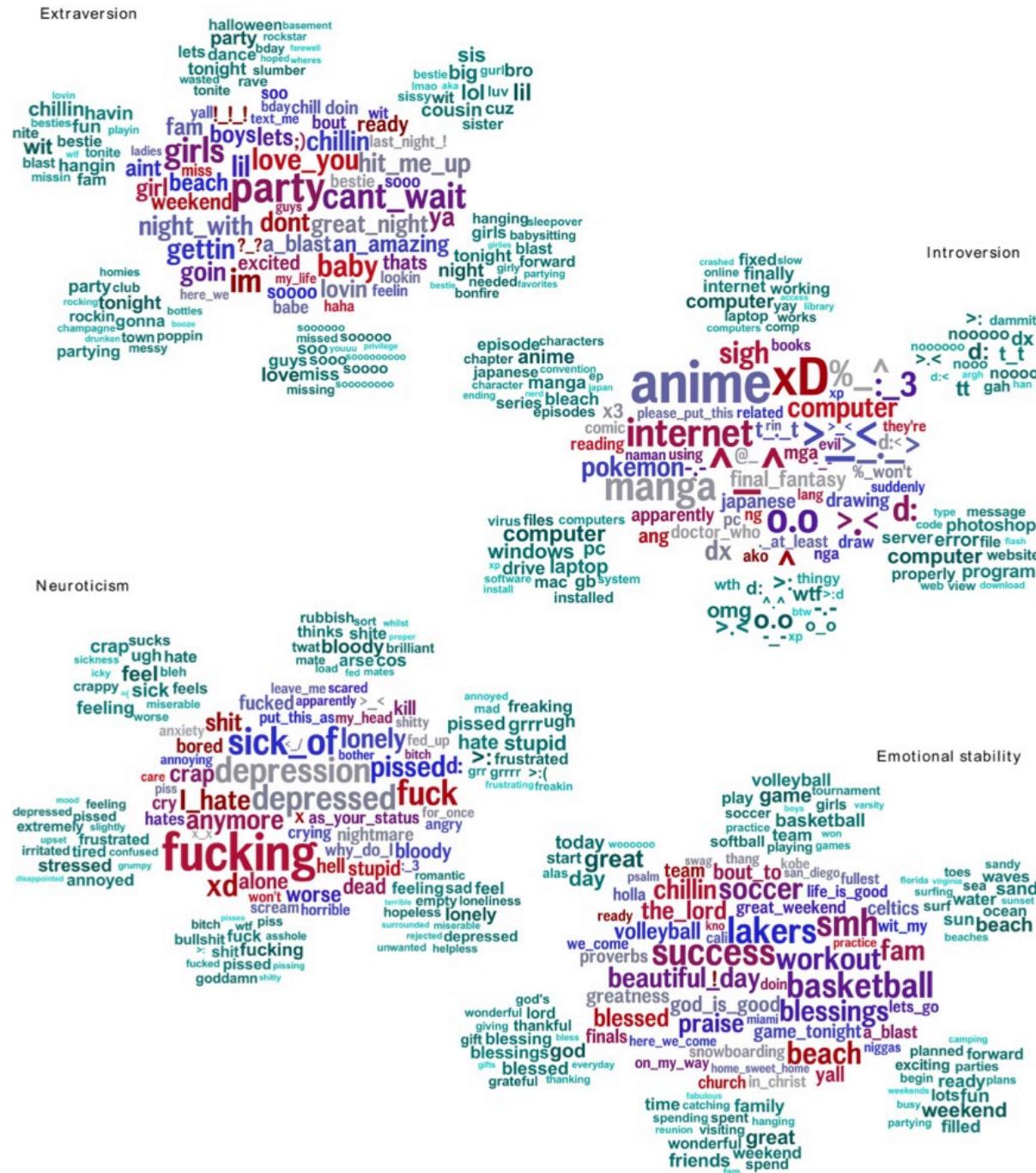
# Python for Facebook

Facebook的Python SDK詳細介紹可以看[這裡](#)

```
import json
def show(j_in):
    print(json.dumps(j_in,ensure_ascii=False,indent=2))

import facebook
token='...' # replaced by your access token
graph=facebook.GraphAPI(token)
show(graph.get_object('me',fields='id,name,email'))
show(graph.get_connections('me','friends'))
#graph.put_object('me','feed',message='Hello from class!')
```

# Twitter上的用語亦可預測性格



# 和憂鬱症相關的語彙

## Depression-indicative posts

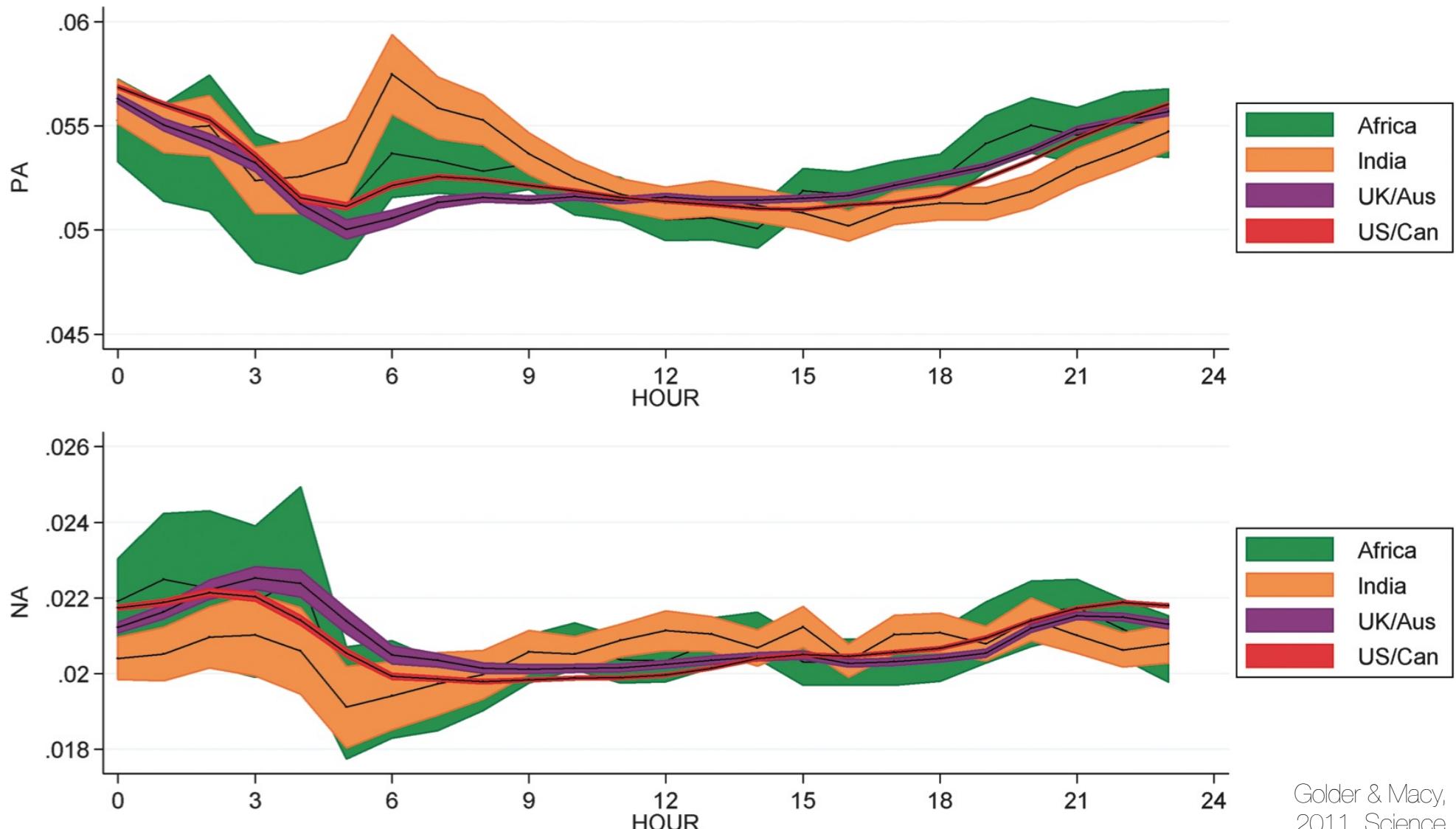
loser, depress\*, lonely, sad, alone, weak, useless, life, imbalance, blame, problems, unsuccessful, suicidal, torture, safe, escape, worry, intimidat\*, uncomfortable, therapy, medication, shit, pressure, conversation, hurts, myself, worth, break, nobody, mine, painful, hate, suck\*

## Standard posts

lol, work, weekend, say, friends, brilliant, follow, tips, bieber, love, amazing, hello, now, bored, awesome, beautiful, romantic, fuck\*, perfect, excited, smile, meet, tonight, life, movie, football, favorite, sleepy, great, night, team, good, anyone, you, your, tomorrow, money

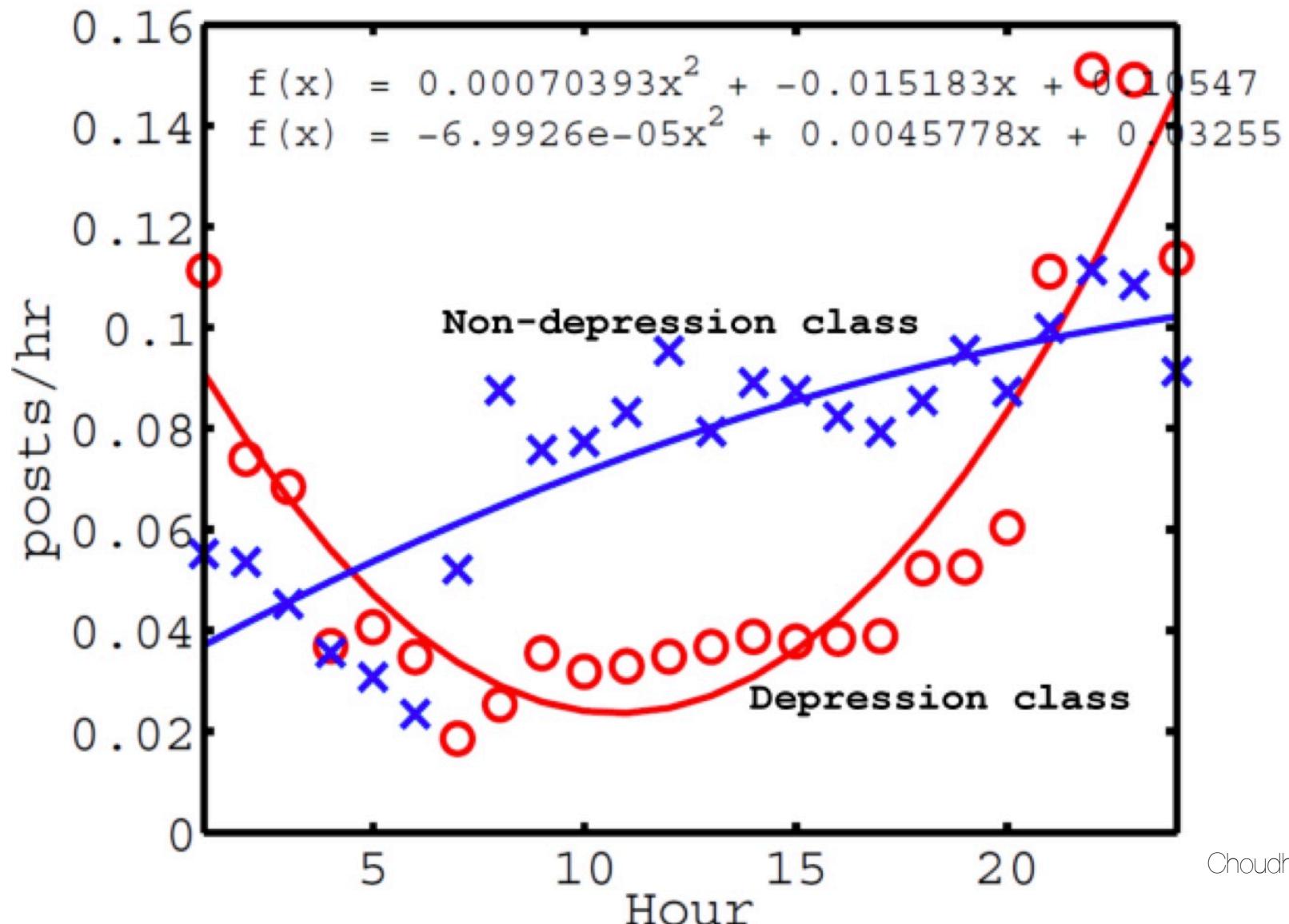
# Twitter上的情緒詞

情緒隨著時間變化



# Twitter上的文章量

也隨著族群不同有所變化



# 需求=f(人,時間)



**Abdulrahman** @BoFhaidan · May 29

Saw this at [@TheSchoolOfLife](#), interesting perspective 🤝 #insomnia



25

29

...

[View photo](#)



**Barnacules** @Barnacules · May 29

Good morning internet... I'm not even sure if I actually went to sleep or not?  
#Insomnia #UploadComingSoon



2

59

...



**AKilluminati** @An0nKn0wledge · May 29

Do i ever sleep?  
Its a short awnser, No..  
#Insomnia #Nightcr3w



2

6

...



**Adam Chang** @AdamChang23 · May 29

I envy people that can fall asleep in like 30 seconds.. 😴 😴 😴 #Insomnia



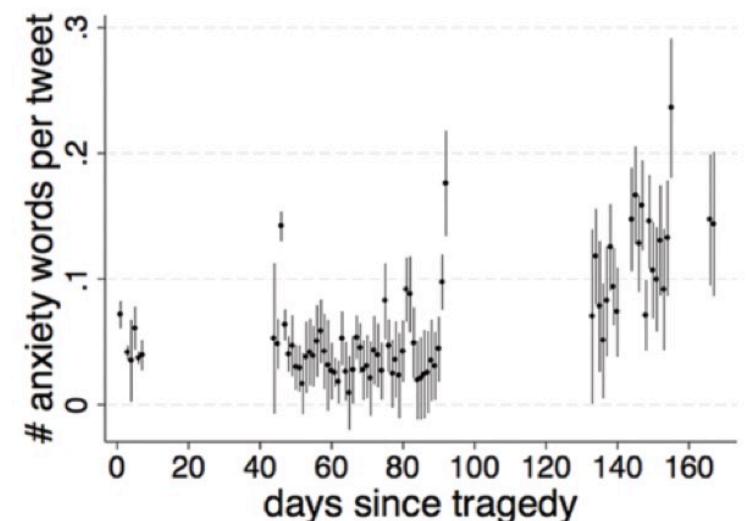
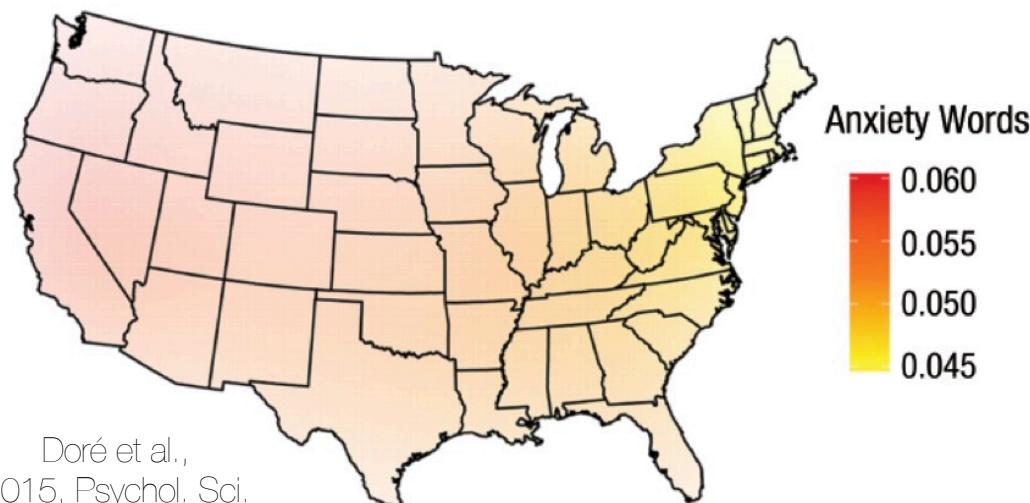
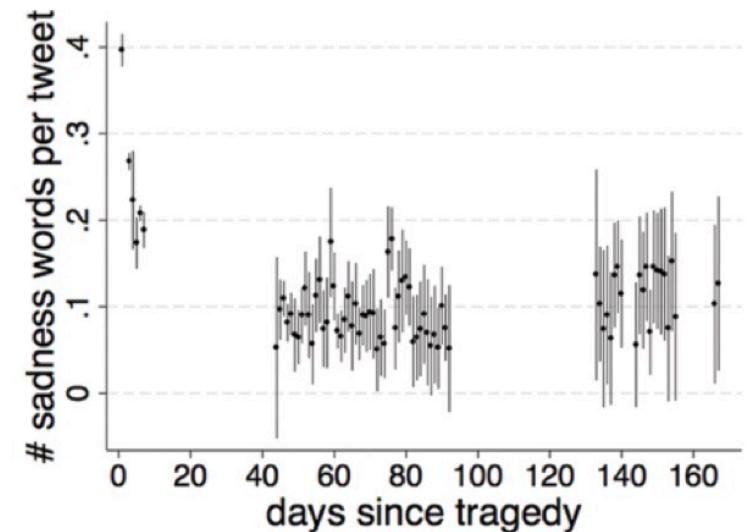
4

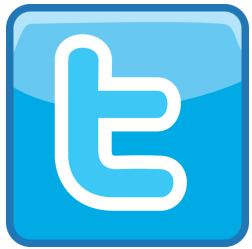
8

...

# 需求=f(人,時間,地點)

煩惱未來導致焦慮





# Twitter API

Twitter APIs的詳細介紹可以看[這裡](#)

踹看看：

<https://twitter.com/search?q=台大>

<https://api.twitter.com/1.1/search/tweets.json?q=台大>

Twitter APIs幾乎都要通過OAuth身份認證才能用

## Application Settings

*Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.*

Consumer Key (API Key) HDTBqZDAOuZJyDhsr5Z5tw

Consumer Secret (API Secret) nKULI80bwwedY4jtk4UTt2dXotoIVOIPQ8fUVK2LDo

## Your access token

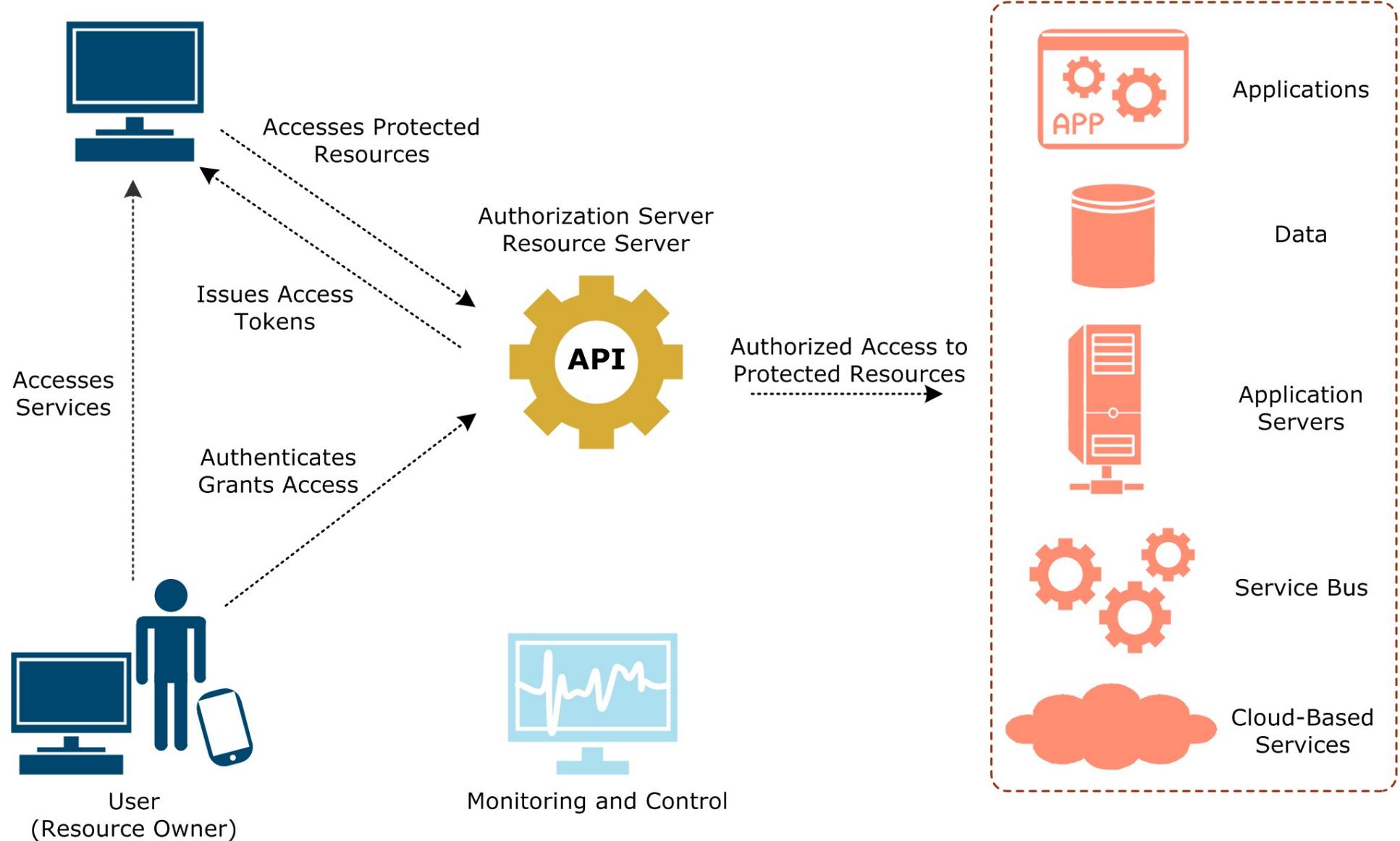
*This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.*

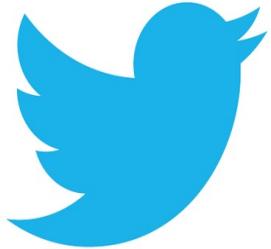
Access token 206228745-LujHvkviaSuBq4FH2Unx92xvYFCBYtoArQknciLF

Access token secret AbuFKDs8oWSFMcbwUOq9D4E0Blx4dh8tfzvQ93Utndvjp

# OAuth 2.0

確保A能不知我們在B的**密碼**卻能存取B的**部分**資訊





# Python for Twitter

有很多非官方的Python模組可以用  
以下按照熱門程度高至低排列

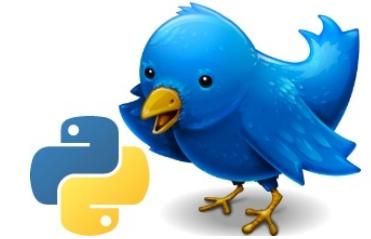
<https://github.com/tweepy/tweepy>

<https://github.com/bear/python-twitter>

<https://github.com/sixohsix/twitter>

更多Tweepy的文件可以看這裡

# Tweepy , 你好!



```
import tweepy
key1='...' # API key
secret1='...' # API secret
key2='...' # Access token
secret2='...' # Token secret
auth=tweepy.OAuthHandler(key1,secret1)
auth.set_access_token(key2,secret2)
api=tweepy.API(auth)
data=api.search(q=u'台灣',
                 geocode='22.9500,120.2000,500km')
for tweet in data:
    print(tweet.user.screen_name,tweet.text,'\n')
# for follower in api.followers_ids('iingwen'):
#     user=api.get_user(follower)
```

# Google有各種服務的APIs

使用前要先創立Projects

The screenshot shows the Google APIs Project management interface. On the left, there's a sidebar with icons for IAM & Admin, All projects (which is selected and highlighted in blue), IAM, Quotas, Service accounts, Labels, Settings, Encryption Keys, and Identity-Aware Proxy. The main area has tabs for Projects, DELETE PROJECT, and HIDE INFO. A prominent red box highlights the '+ CREATE PROJECT' button. Below it, a section titled 'Select a project' contains 'PERMISSIONS' and 'LABELS' tabs, with a message stating 'Please select at least one resource.' A search bar is at the top, and a navigation bar with various icons is at the very top.

# API Keys

使用APIs前要先取得授權碼(API key)

The screenshot shows the Google API Manager interface. The top navigation bar includes the Google logo, the project name "Informatics" (with a dropdown arrow), a search bar, and several icons. The left sidebar has sections for "API Manager", "Dashboard", "Library", and "Credentials", with "Credentials" currently selected and highlighted by a red box. The main content area is titled "Credentials" and contains three tabs: "Credentials" (selected), "OAuth consent screen", and "Domain verification". A large callout box highlights the "API key" section, which is described as "Identifies your project using a simple API key to check quota and access". Below it are descriptions for "OAuth client ID" and "Service account key". At the bottom of the callout box is a blue button labeled "Create credentials".

# Google APIs

還要記得enable對應的服務

Google APIs Informatics ▾

API Manager Library

Dashboard

Library

Credentials

Social APIs

Google+ API

Blogger API

Google+ Pages API

Google+ Domains API

YouTube APIs

YouTube Data API

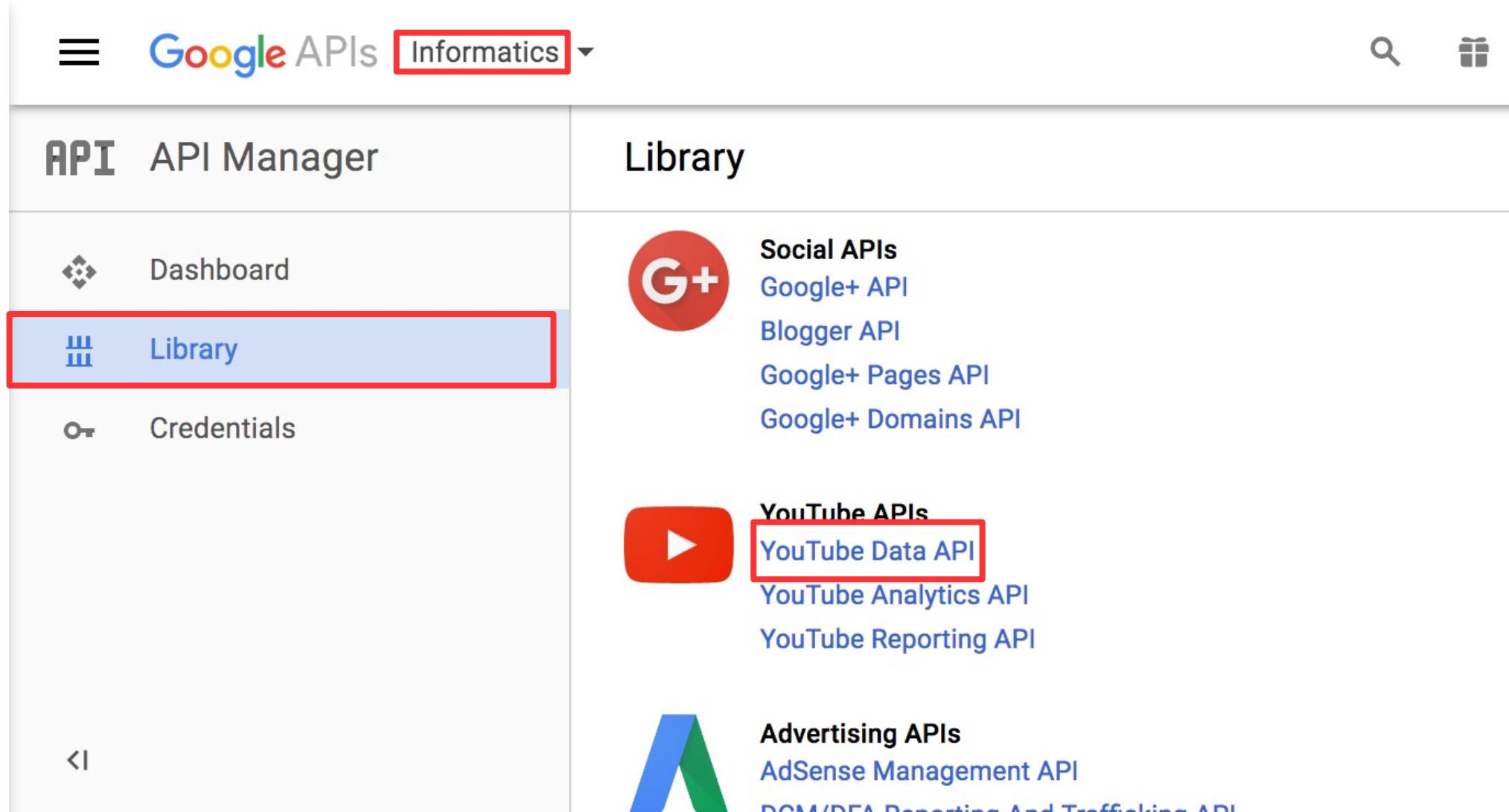
YouTube Analytics API

YouTube Reporting API

Advertising APIs

AdSense Management API

DCM/DEA Reporting And Trafficking API

The screenshot shows the Google APIs console interface. On the left, there's a sidebar with three main items: 'Dashboard', 'Library' (which is highlighted with a red border), and 'Credentials'. The main area is titled 'Library' and contains sections for 'Social APIs' (listing Google+, Blogger, Google+ Pages, and Google+ Domains APIs) and 'YouTube APIs' (listing YouTube Data, Analytics, and Reporting APIs). At the bottom, there's another section for 'Advertising APIs' (listing AdSense Management and DCM/DEA APIs). A red box highlights the 'Library' item in the sidebar and the 'YouTube Data API' item under the YouTube APIs section.

# Python for YouTube

發燒影片 #1

【震撼】一切都不是巧合，2020真正的預言 | 老高與小茉 Mr & Mrs Gao

觀看次數：1,848,795次

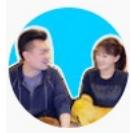
5.6  
萬

1127

分享

儲存

...



老高與小茉 Mr & Mrs Gao ✓  
257萬 位訂閱者

加入

訂閱

【加入會員按鈕】<https://www.youtube.com/channel/UCMU...>

【出賣靈魂按鈕】<https://goo.gl/VhzZeS>

顯示完整資訊

```
key="..." # Developer key
from apiclient.discovery import build
youtube = build('youtube', 'v3', developerKey=key)
result = youtube.videos().list(
    part='snippet,contentDetails,statistics',
    chart='mostPopular',
    regionCode='TW',maxResults=50).execute()
print(result)
```

# 其他社群媒體的APIs

請自己花時間探索

Line

Telegram

Instagram

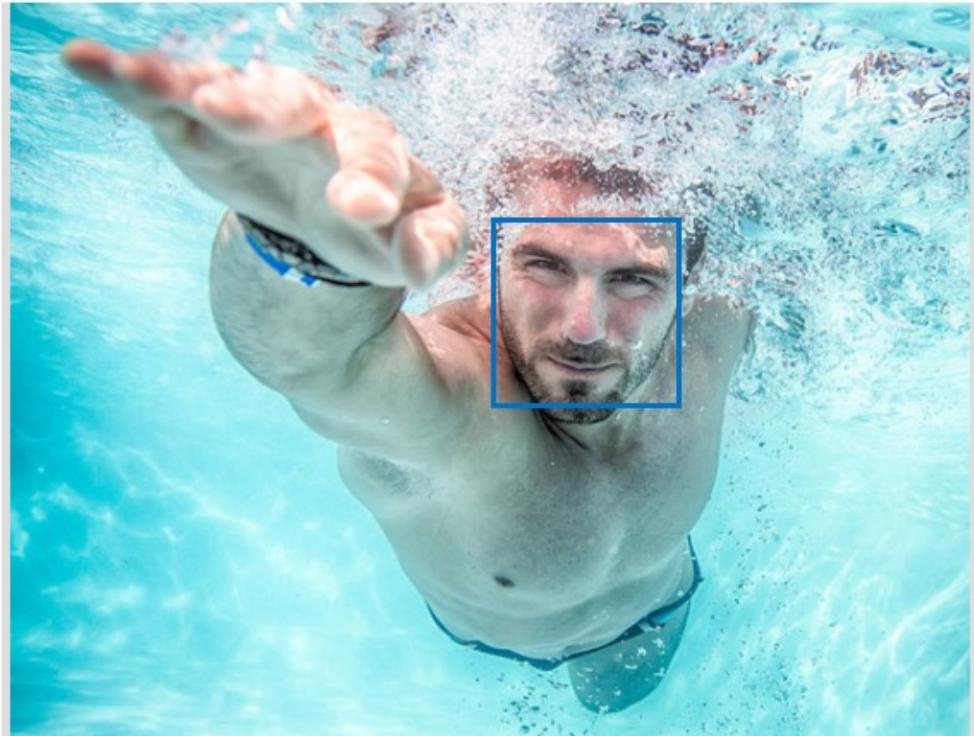
...



若沒提供則用下週方法

# Beyond Social Media

有很多其他類型的資料庫或應用程式APIs



FEATURE	VALUE
NAME:	
Description	{ "tags": [ "water", "swimming", "sport", "pool", "person", "man", "frisbee", "ocean", "blue", "bird", "riding", "top", "standing", "wave", "young", "body", "large", "game", "glass", "pond", "playing", "board", "catch", "clear", "boat", "white" ], "captions": [ { "text": "a man swimming in a pool of water", "confidence": 0.8909298 } ] }
Tags	[ { "name": "water", "confidence": 0.9997857 }, { "name": "swimming", "confidence": 0.955619633 }, { "name": "sport", "confidence": 0.953807831 }, { "name": "pool", "confidence": 0.9515978 }, { "name": "person", "confidence": 0.889862537 }, { "name": "water sport", "confidence": 0.664259 } ]
Image format	"Jpeg"

例如Microsoft/Google的Computer Vision APIs

# 本週作業

進一步研究Social Media & Web APIs

1. 使用YouTube APIs研究熱門影片的特徵  
(如片長、主題、影片內容等)

(可以是data- or hypothesis-driven research)

2. 精簡的分析報告中至少要兩張統計圖或表來支持論述  
(一張4分;不算影片範例圖)

請在下週一0am前繳交Jupyter NB輸出的.html檔

# Game Over

