

Python基礎 (爬蟲+一點點探勘)

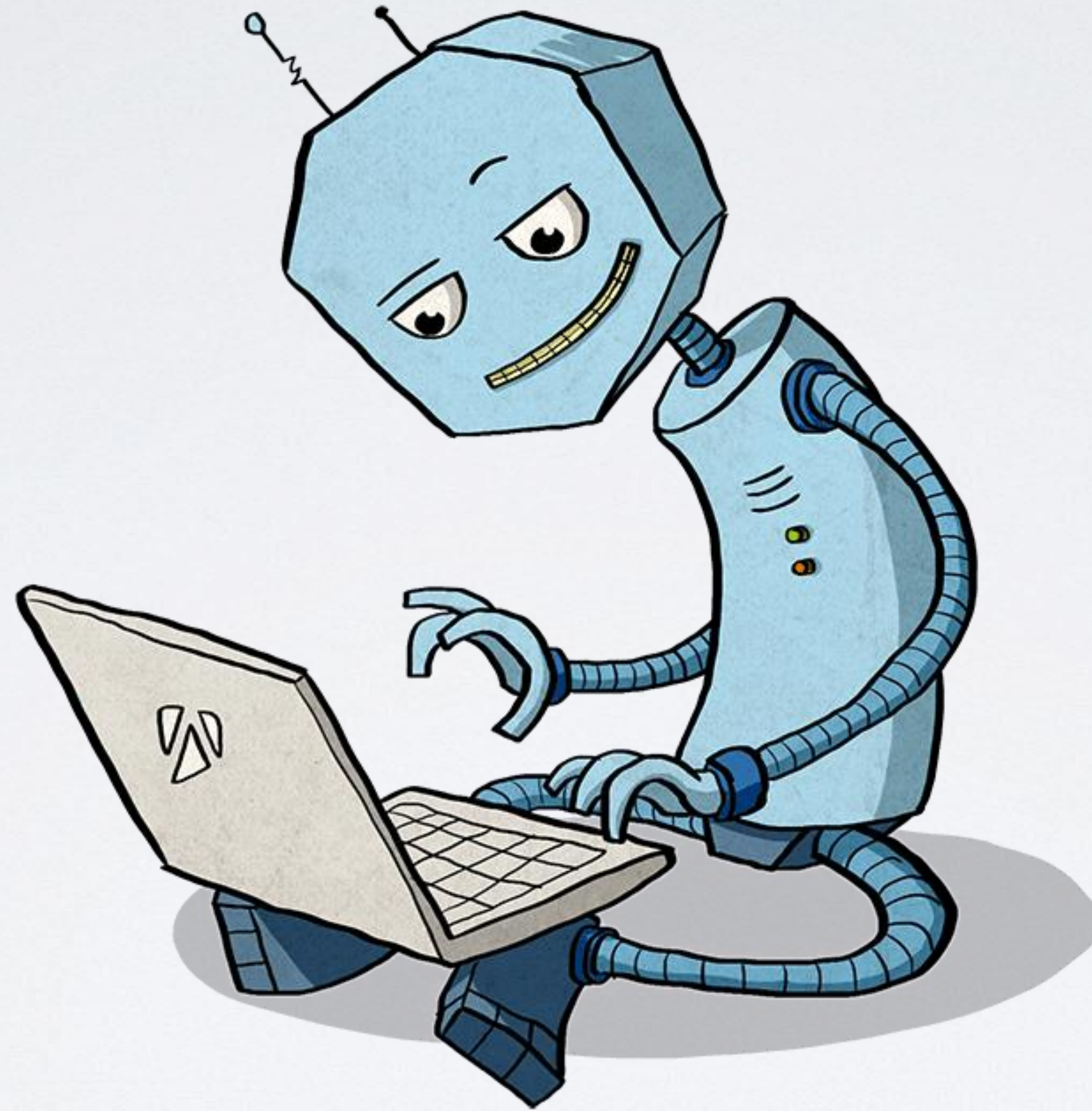
洪培仁 Tim Hong

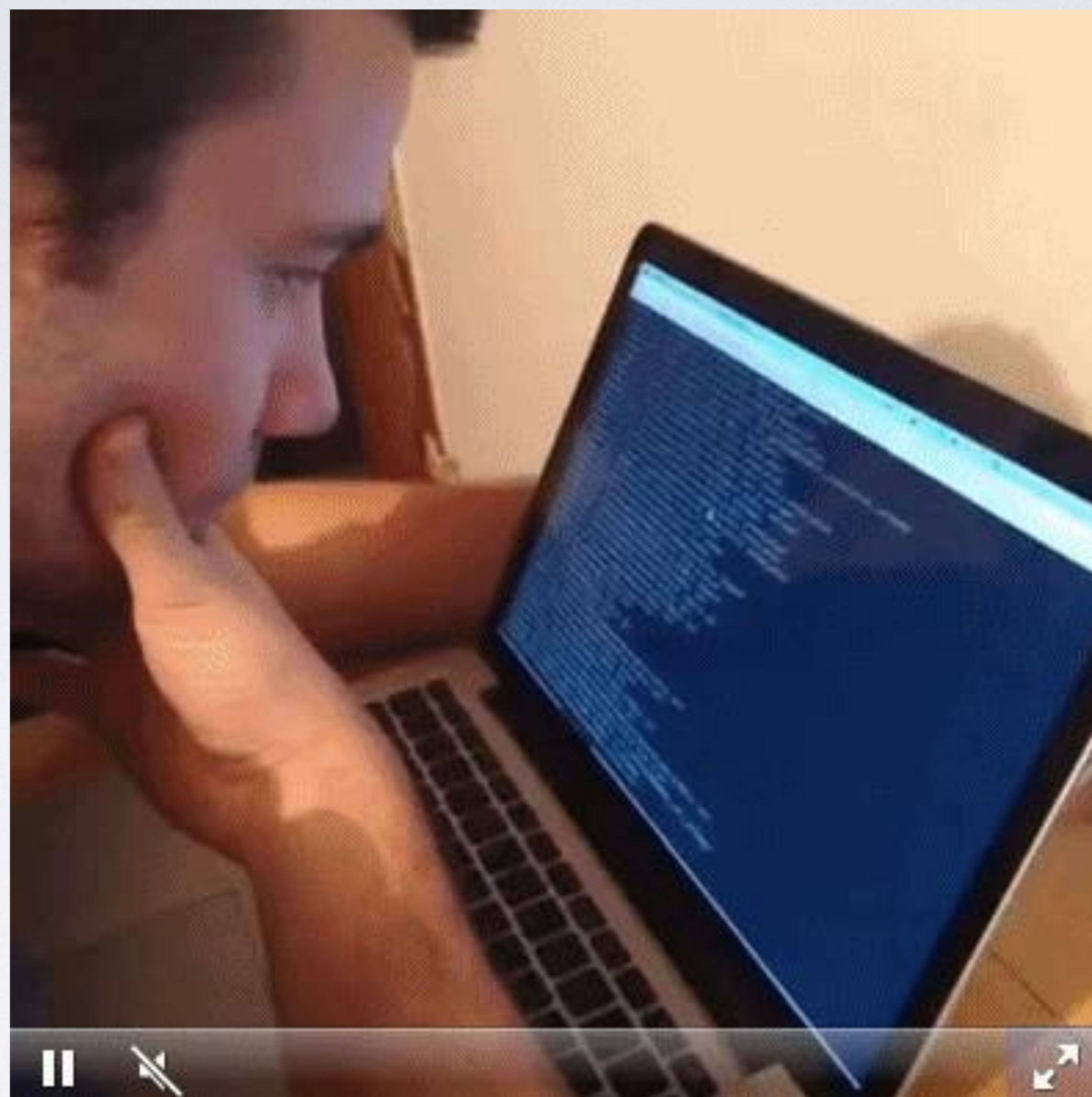
應用程式、資料工程

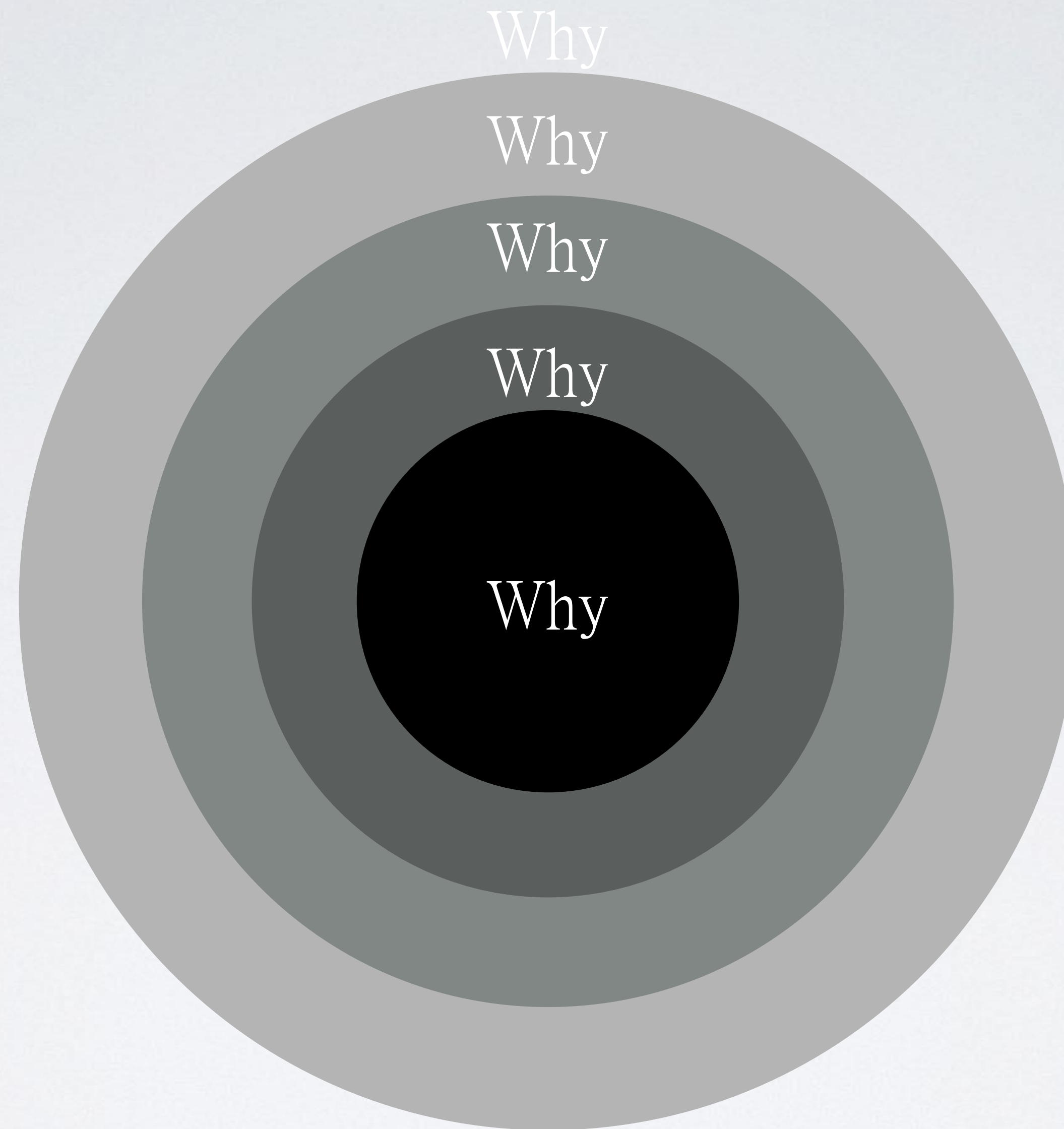
2 Startup、1 NPO、2 Apps 1st

- BenQ Prototype Developer
 - DSP智庫驅動 應用顧問
 - 台北市政府 公參會 OpenData 召集人
-
- 企業應用顧問
 - 中華科技大學 業界講師
 - 波意設計、深圳波意科技
 - 上擎科技
 - 華碩電腦

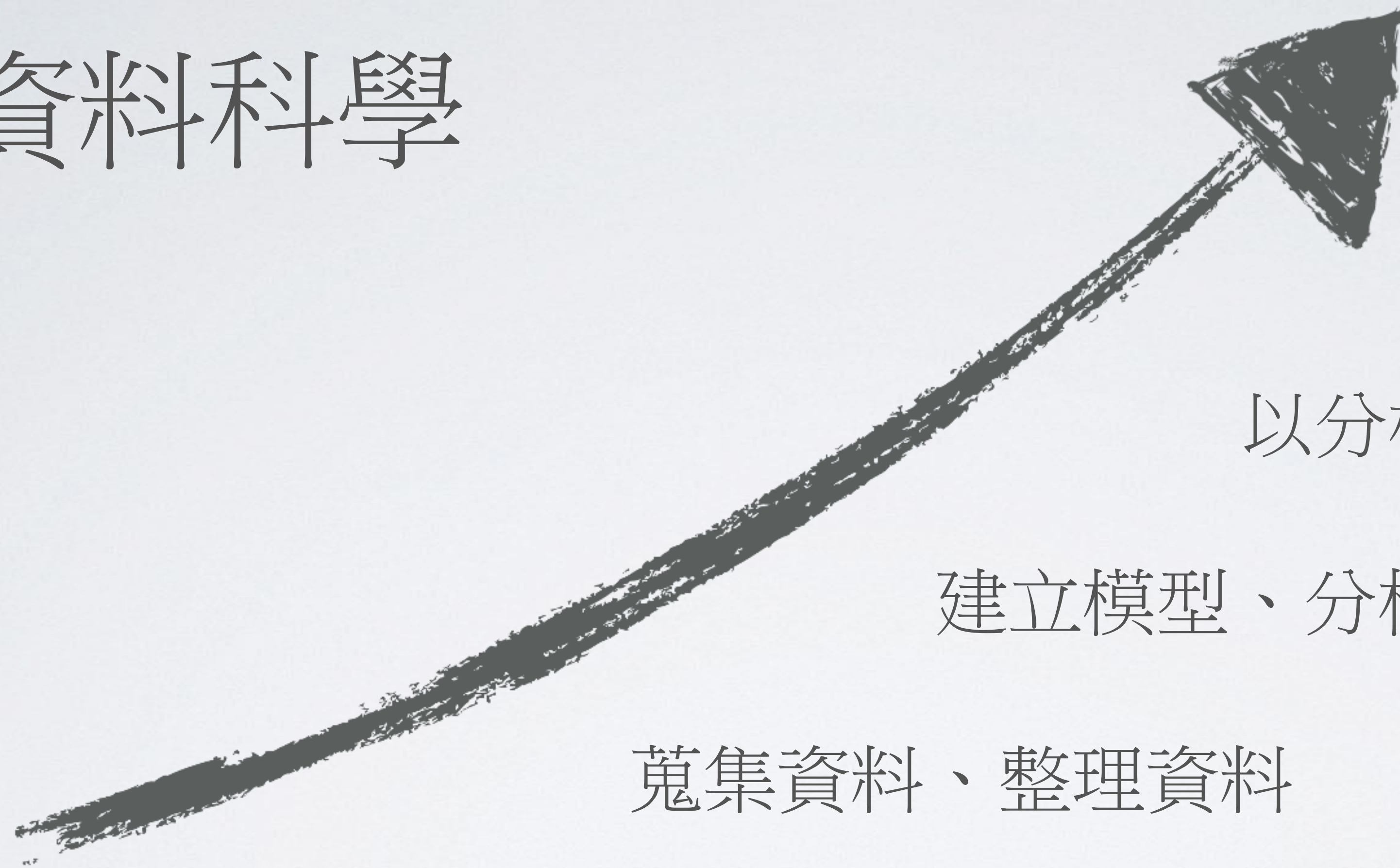








資料科學



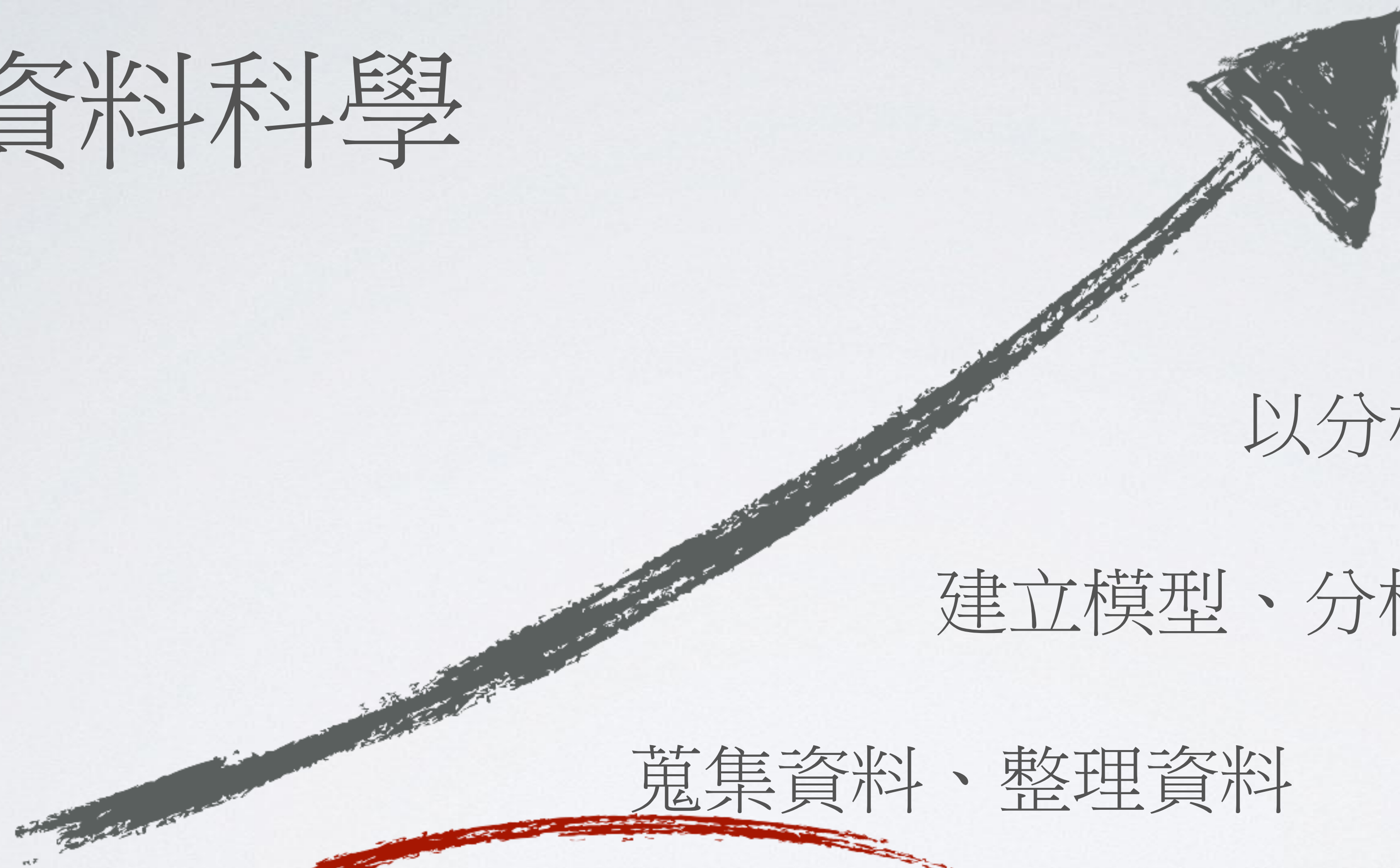
以分析結果回答問題

建立模型、分析資料

蒐集資料、整理資料

定義問題、規劃藍圖

資料科學



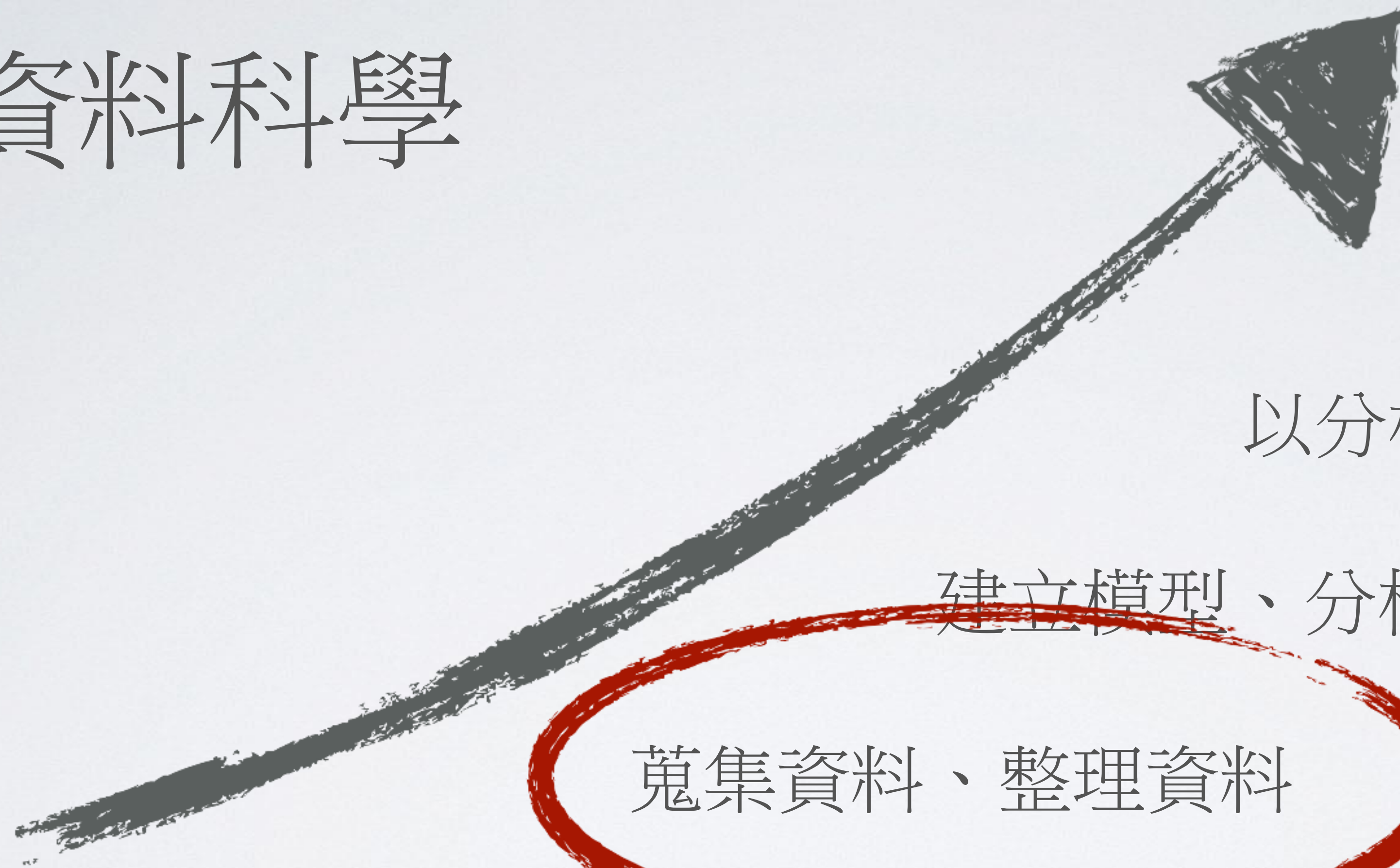
以分析結果回答問題

建立模型、分析資料

蒐集資料、整理資料

定義問題、規劃藍圖

資料科學



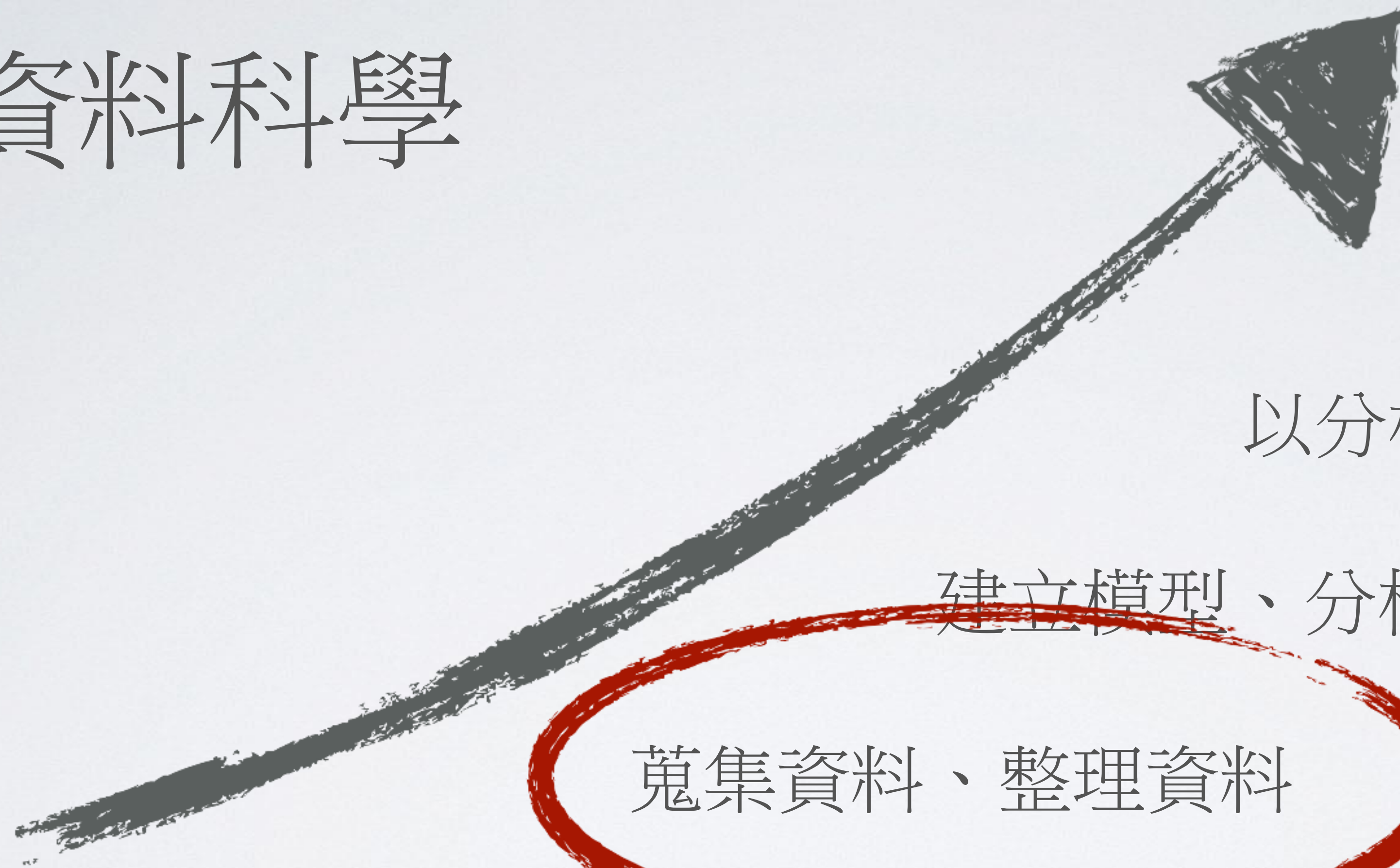
以分析結果回答問題

建立模型、分析資料

蒐集資料、整理資料

定義問題、規劃藍圖

資料科學



以分析結果回答問題

建立模型、分析資料

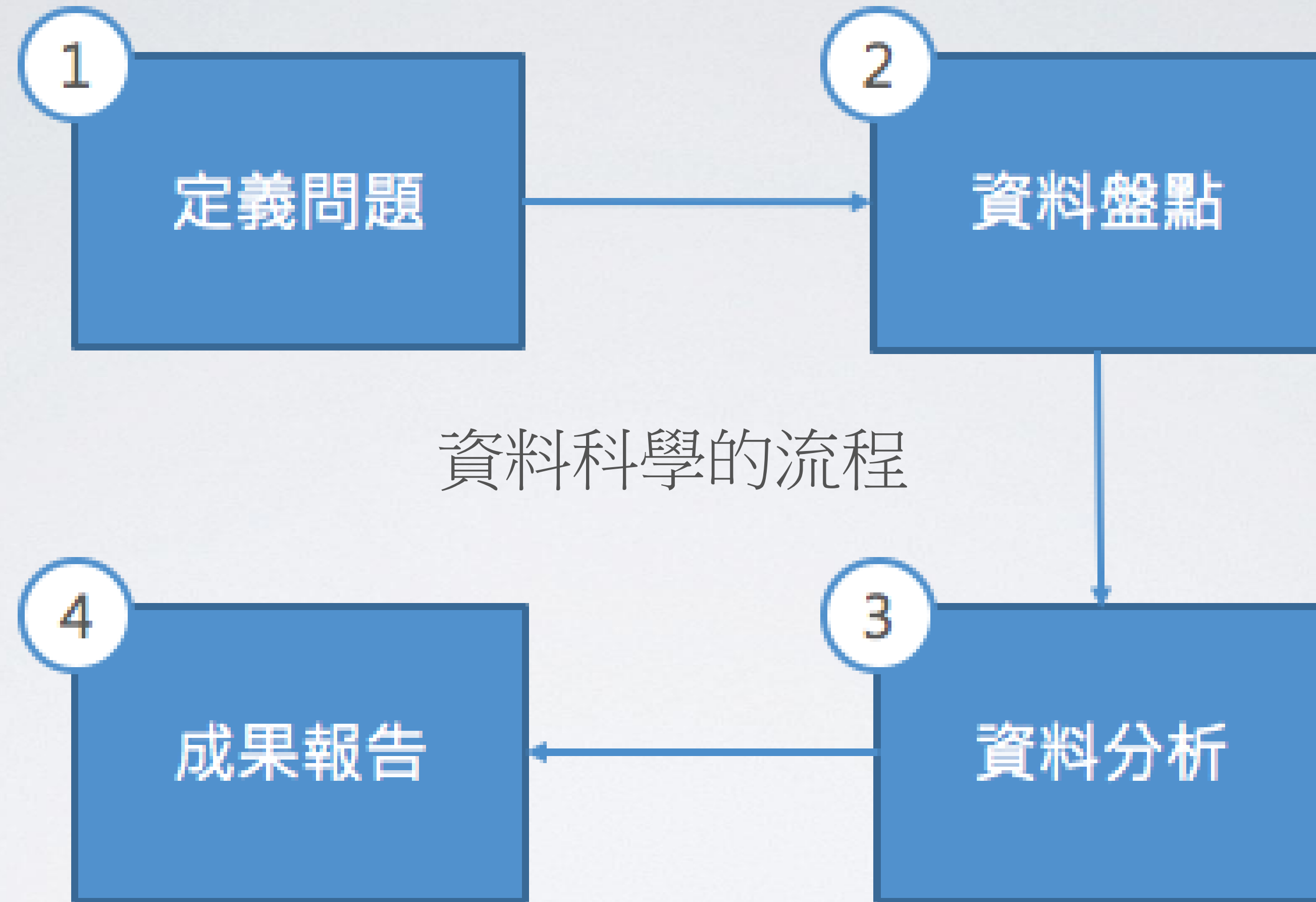
蒐集資料、整理資料

定義問題、規劃監圖

80%

資料的品質影響深遠









Guido Rossum



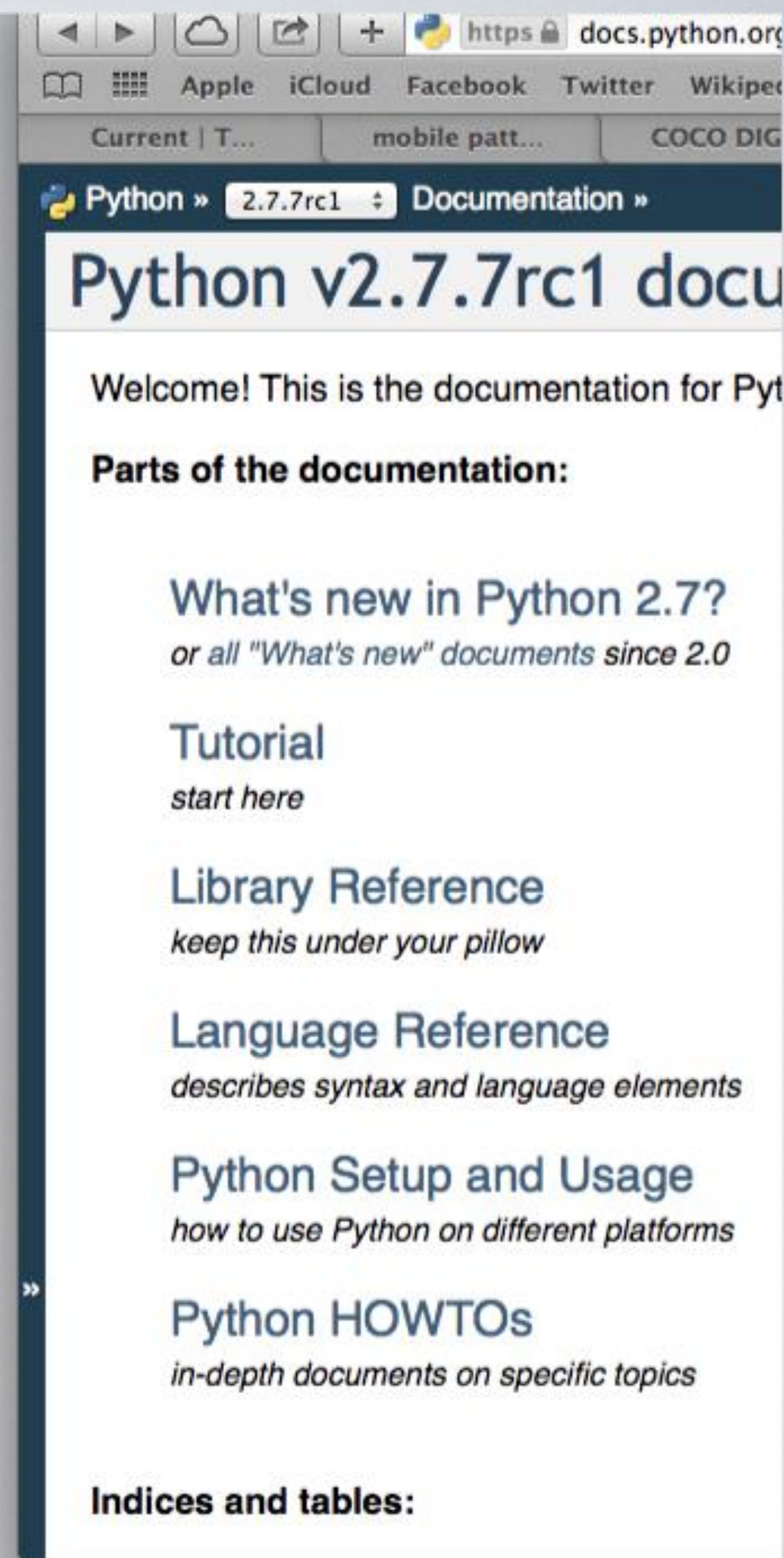
使用哪一個版本？

```
[Tim-MacBook-Pro-5:~ tim$ python
Python 2.7.11 (v2.7.11:6d1b6a68f775, Dec  5 2015, 12:54:16)
[GCC 4.2.1 (Apple Inc. build 5666) (dot 3)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> █
```

2.7 or 3 都很好
不過有些命名的差異需要注意

相關資訊在哪

- <https://docs.python.org/2/>
- 建議讀過所有文件
- 進行官方教學

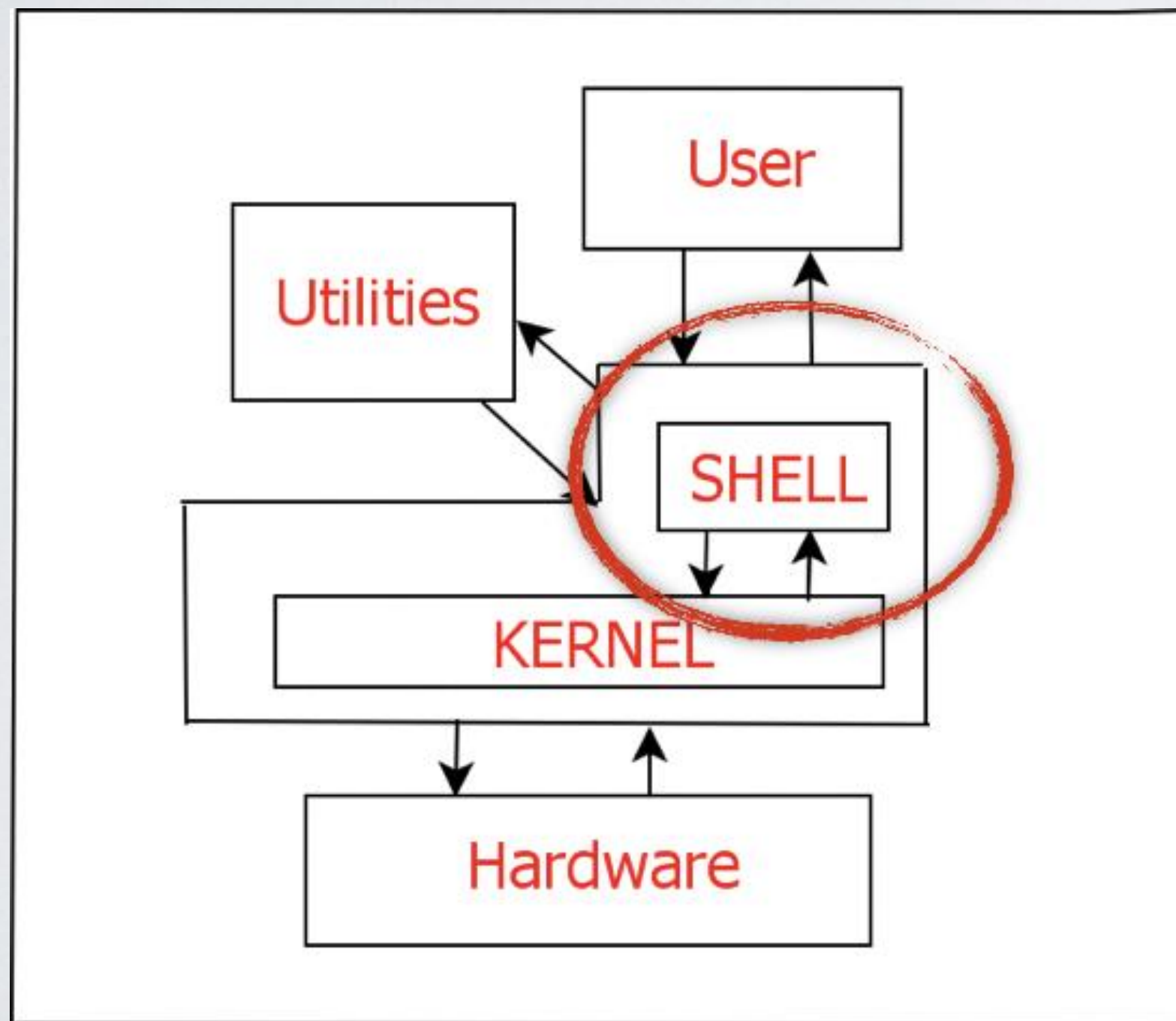


How to do it?

Type python in your terminal or cmd
Linux

Type in python2 or python3

啟動Python Shell



It should looks like this.

```
[Tim-MacBook-Pro-5:~ tim$ python
Python 2.7.11 (v2.7.11:6d1b6a68f775, Dec  5 2015, 12:54:16)
[GCC 4.2.1 (Apple Inc. build 5666) (dot 3)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>>
```




開發工具下載位置

← → ↻ 🏠 🔒 Secure | https://www.anaconda.com/download/ ☆ 🗨️ 📺 📄

📱 Apps 📊 GA&R 🐼 Johnny-Five: The JavaScrip 🎧 📺 📄 Shiny Dashboard 🌐 SimilarWeb 📺 Build Web Apps with G 📄 RPubS - basics of regress 📄 HTML Formatter 📄 迴歸

ANACONDA

Anaconda Cloud Documentation Blog Contact 🔍

What is Anaconda? Products Support Community About Resources **Download**

🪟 Windows 🍏 macOS 🐧 Linux

Anaconda 5.0.0 For Windows Installer

Python 3.6 version *
64-Bit Graphical Installer (535 MB) ⓘ

⬇️ **Download**

[Download 32-Bit \(436 MB\)](#)

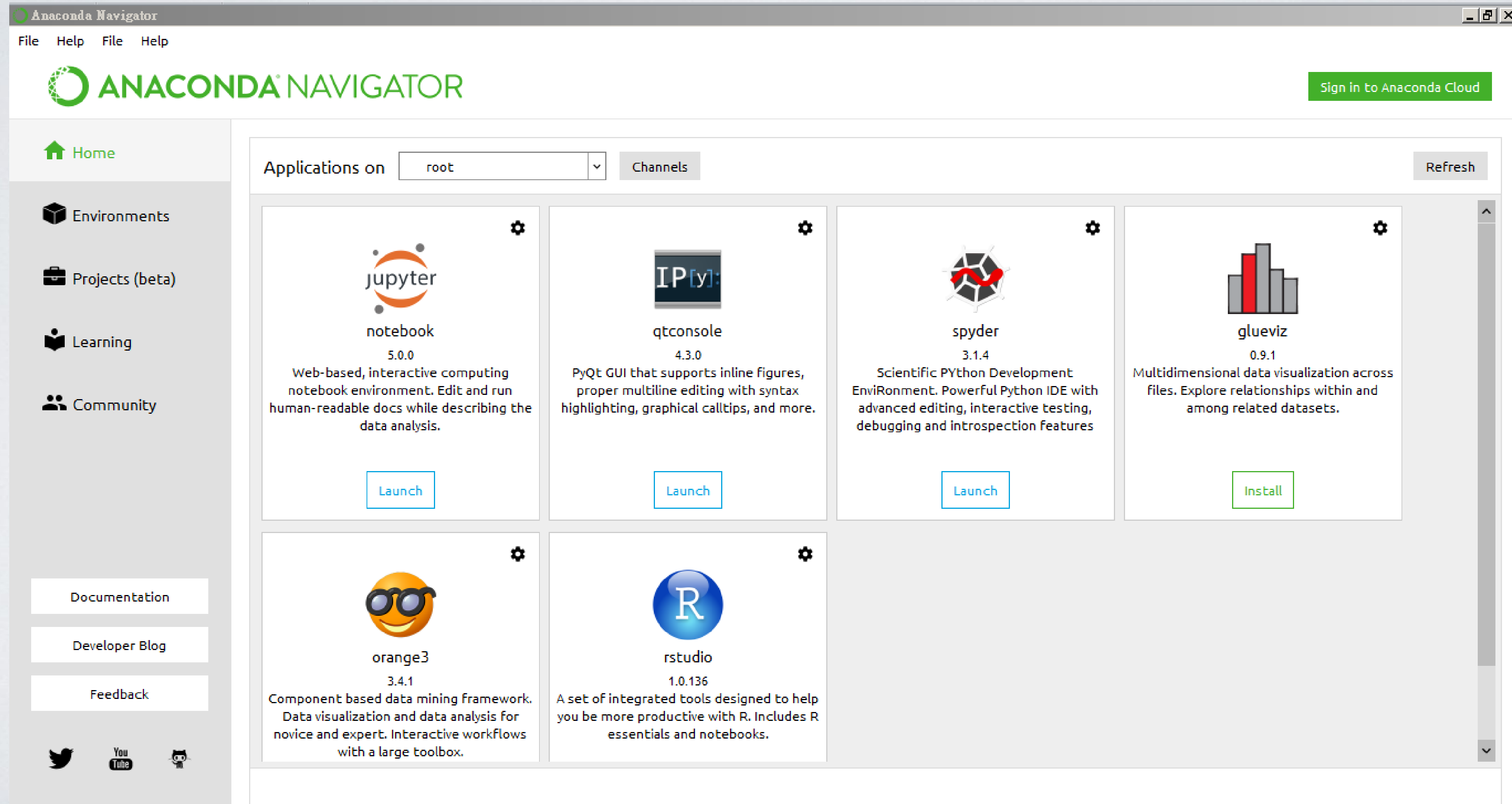
Python 2.7 version *
64-Bit Graphical Installer (522 MB) ⓘ

⬇️ **Download**

[Download 32-Bit \(421 MB\)](#)

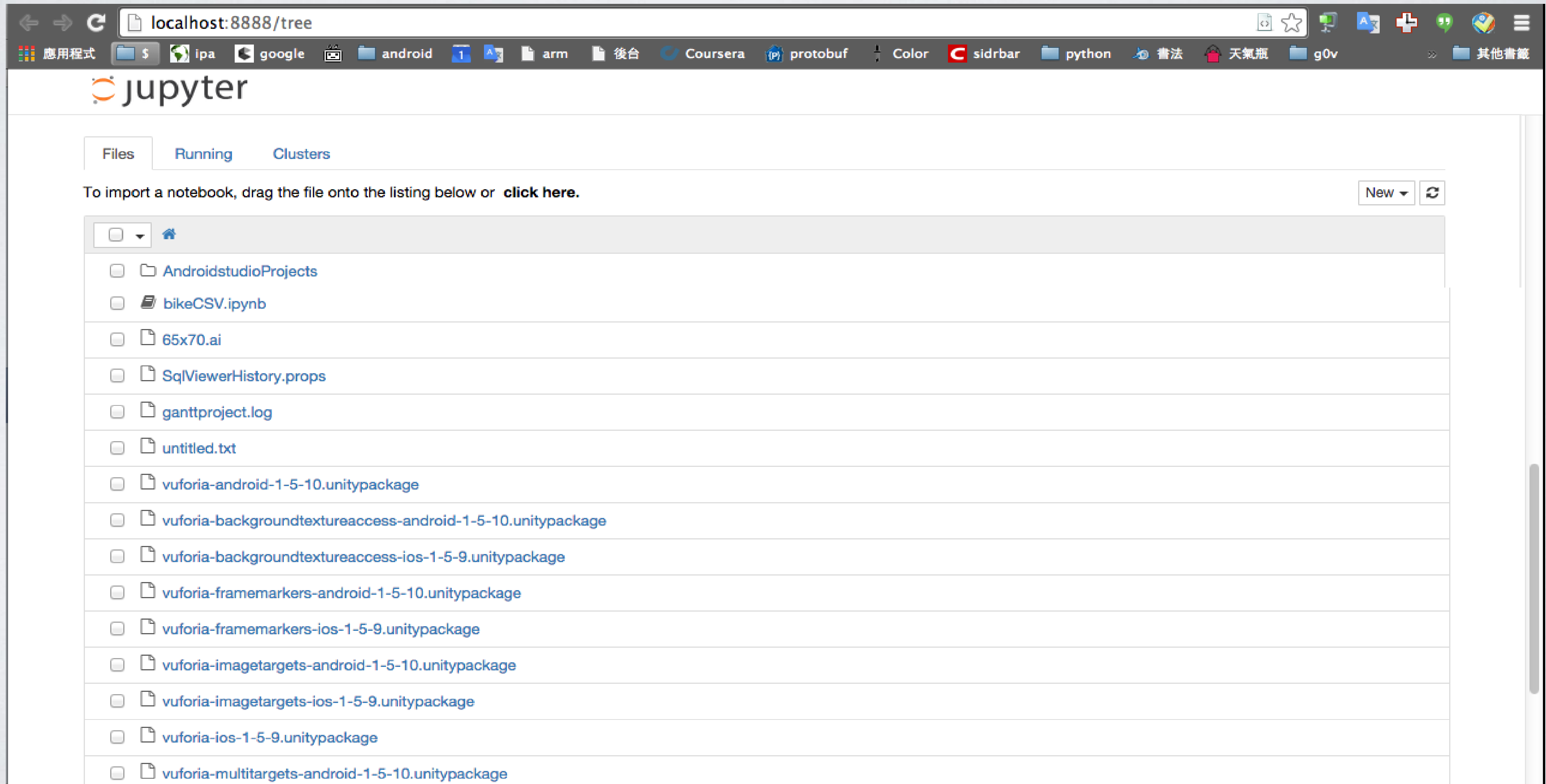
CH 📄 ? ⌵

開發工具



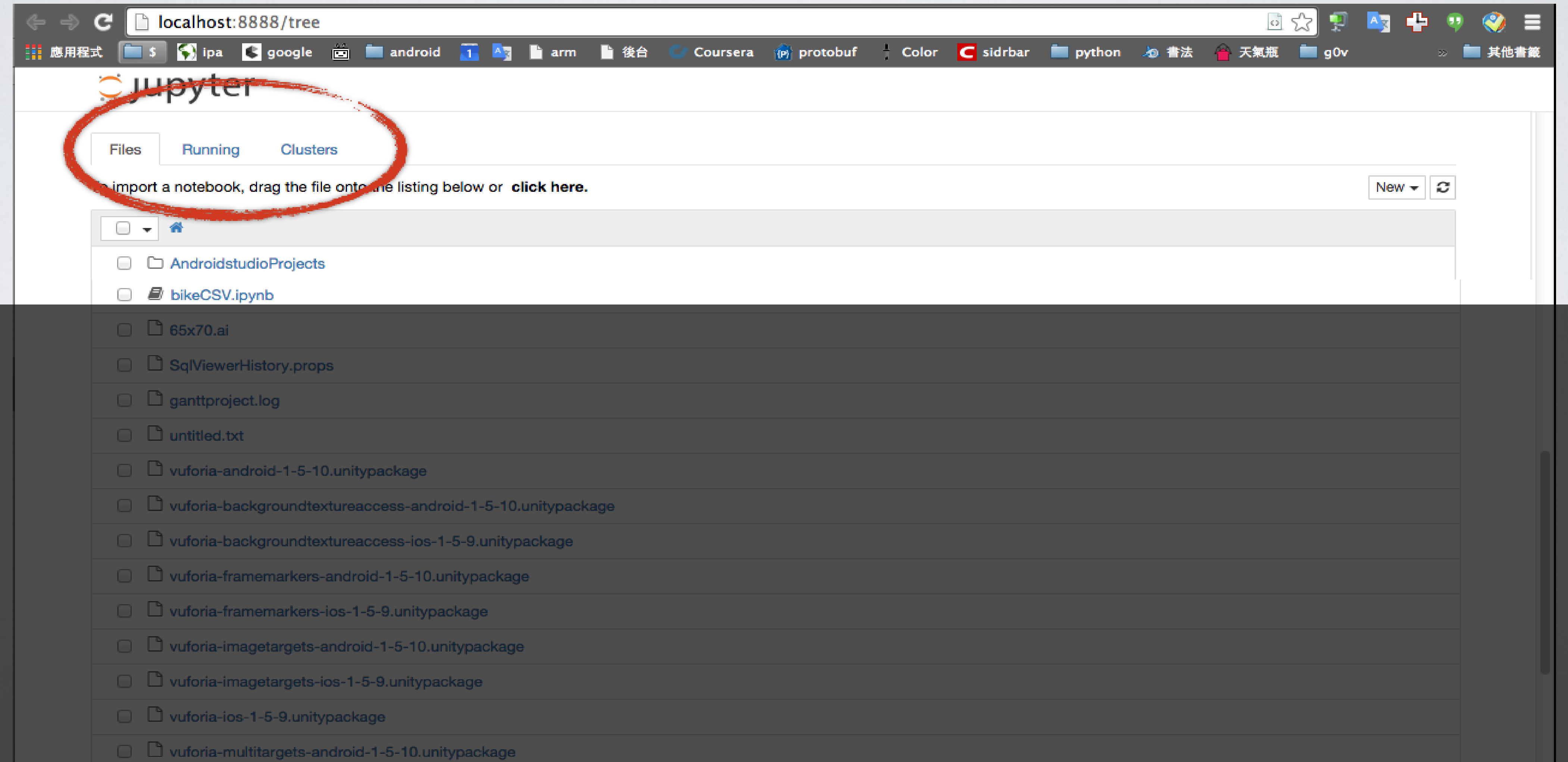
IPYTHON NOTEBOOK介紹1

Web-based 的開發介面



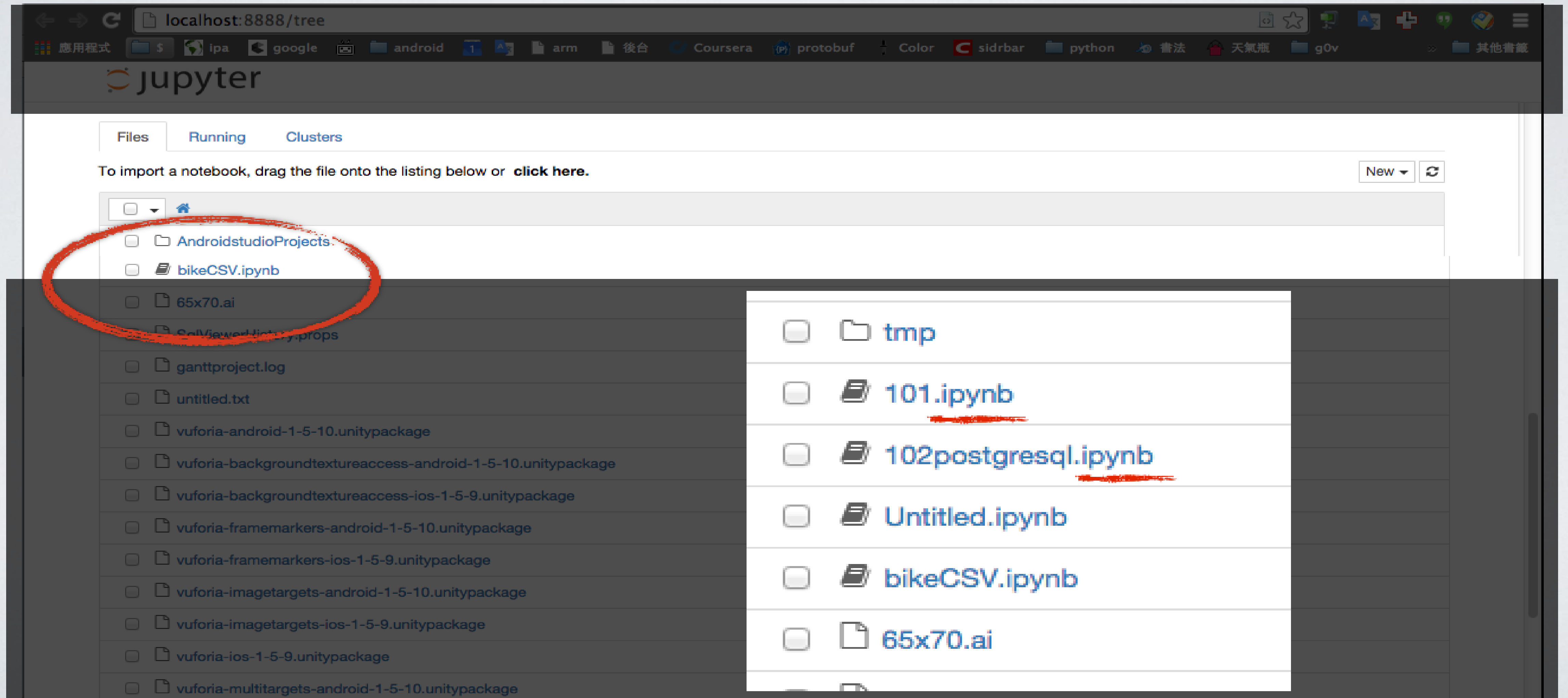
IPYTHON NOTEBOOK介紹II

Server 狀態



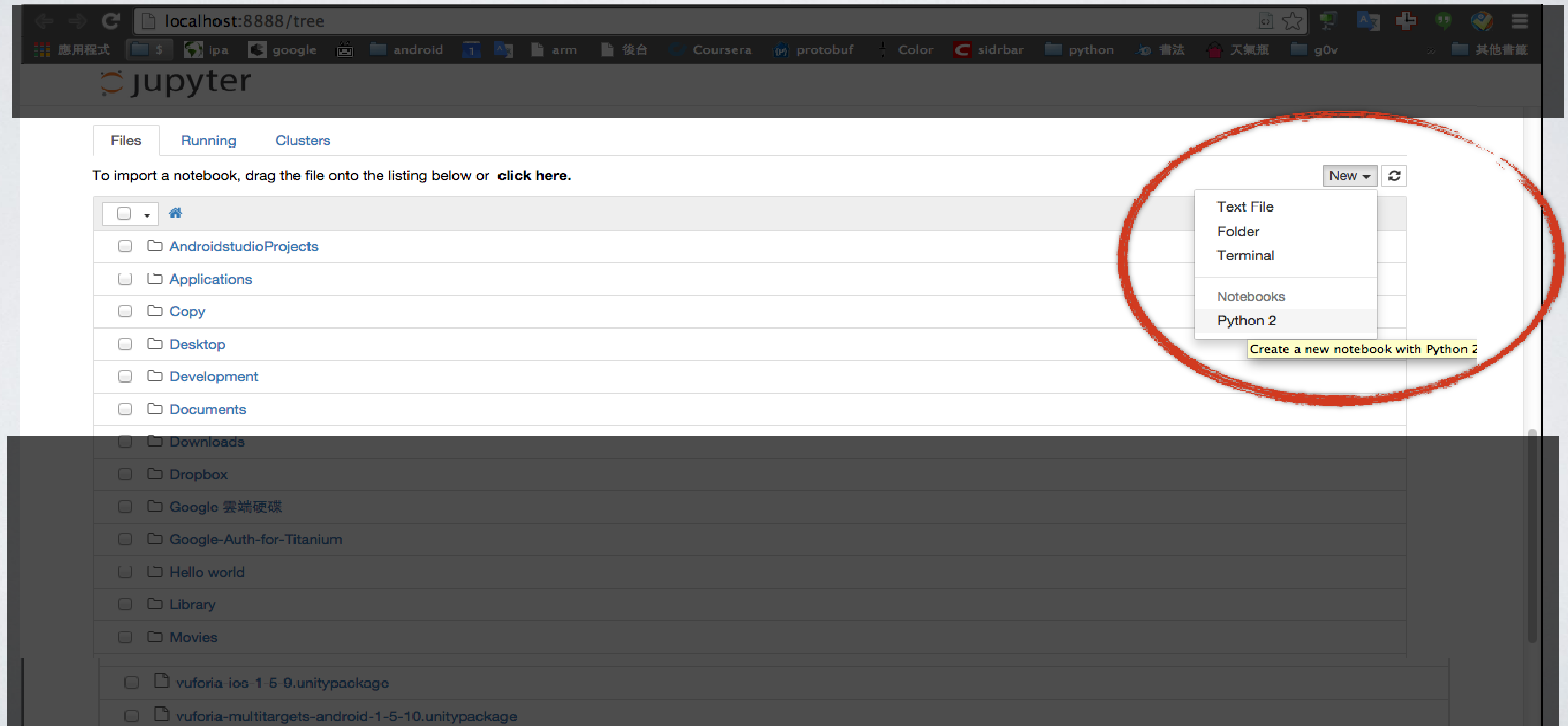
IPYTHON NOTEBOOK介紹III

已經編譯過的程式



IPYTHON NOTEBOOK

Hello World



IPYTHON NOTEBOOK

Hello World

The image shows a Jupyter Notebook interface. At the top, the Jupyter logo is followed by the text "jupyter" and "Untitled1". To the right of "Untitled1" is the text "Last Checkpoint: a few seconds ago (autosaved)". Below this is a menu bar with "File", "Edit", "View", "Insert", "Cell", "Kernel", and "Help". To the right of the menu bar is a status bar that says "Kernel starting, please wait..." followed by a pencil icon and "Python 2". Below the menu bar is a toolbar with icons for saving, adding a new cell, undo, redo, copy, paste, up, down, run, and a cell toolbar dropdown menu set to "None". The main area of the notebook is a dark gray background. A white rectangular box, representing a code cell, is highlighted with a red border. Inside this box, the text "In []:" is followed by a vertical cursor. A red callout box with the word "Cell" inside points to the white box.

jupyter Untitled1 Last Checkpoint: a few seconds ago (autosaved)

File Edit View Insert Cell Kernel Help

Kernel starting, please wait... | Python 2

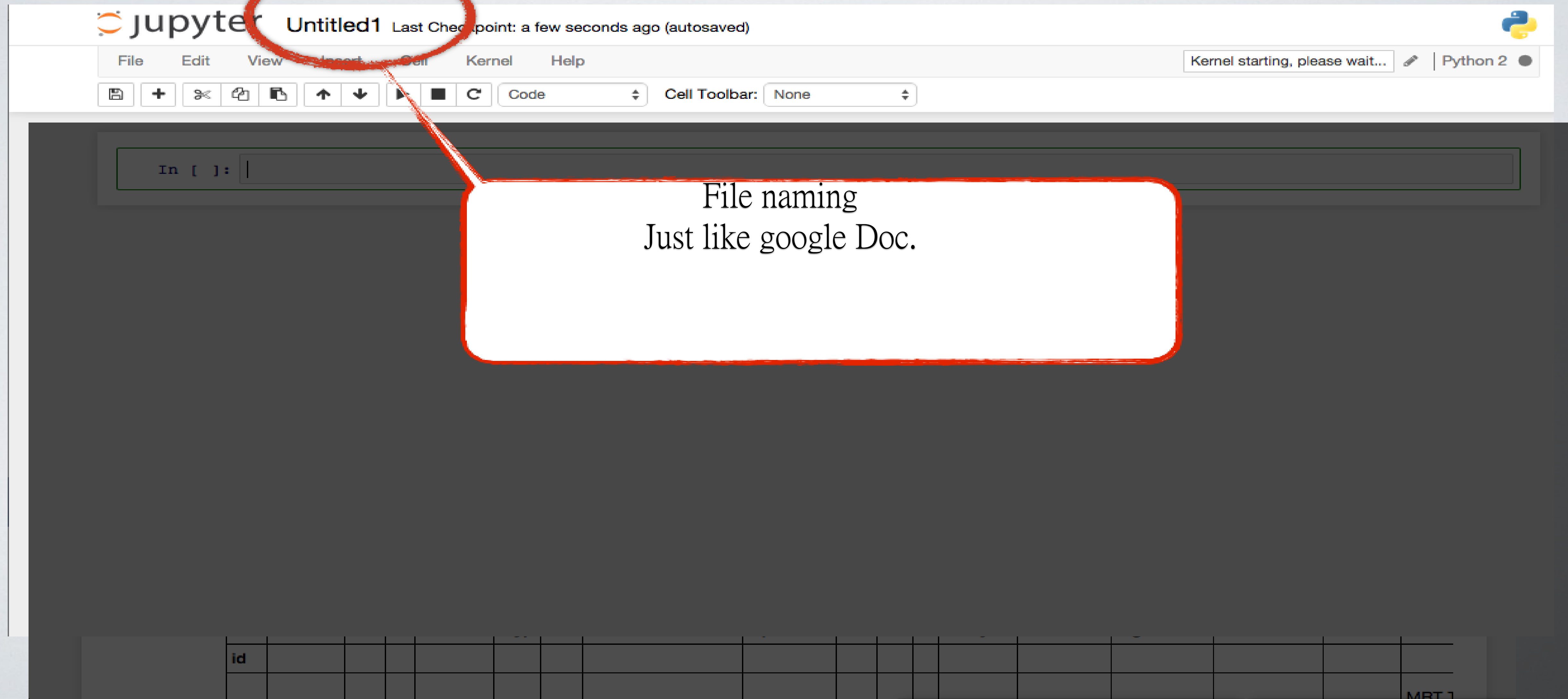
Code Cell Toolbar: None

In []: |

Cell

IPYTHON NOTEBOOK

Hello World



The screenshot shows the Jupyter Notebook web interface. At the top, the Jupyter logo is on the left, and the Python logo is on the right. In the center, the tab is labeled 'Untitled1' with a red circle around it. To the right of the tab, it says 'Last Checkpoint: a few seconds ago (autosaved)'. Below the tab bar is a menu bar with 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', and 'Help'. To the right of the menu bar is a status bar that says 'Kernel starting, please wait...' and 'Python 2'. Below the menu bar is a toolbar with icons for saving, adding, deleting, copying, pasting, undo, redo, and running. To the right of the toolbar is a dropdown menu for 'Code' and a 'Cell Toolbar' dropdown set to 'None'. The main area is a dark gray workspace with a code cell containing 'In []: |'. A red callout box points to the 'Untitled1' tab with the text 'File naming Just like google Doc.'

jupyter Untitled1 Last Checkpoint: a few seconds ago (autosaved)

File Edit View Insert Cell Kernel Help

Kernel starting, please wait... Python 2


Code Cell Toolbar: None

In []: |

File naming
Just like google Doc.

IPYTHON NOTEBOOK

Hello World

 **jupyter** HelloWorld Last Checkpoint: 12 minutes ago (autosaved)

File Edit View Insert Cell Kernel Help

          Code Cell Toolbar: None

```
In [1]: Text1 = 'Hello'
```

```
In [2]: print Text1
```

Hello

```
In [3]: Text2 = 'World'
```

```
In [4]: print Text1+" "+Text2
```

Hello World

```
In [ ]:
```


宣告

```
In [1]: Text1 = 'Hello'
```

```
In [2]: print Text1
```

```
Hello
```

直接宣告 自動判斷

```
In [33]: mylist =["1",2,"AA"]  
mylist
```

```
Out[33]: ['1', 2, 'AA']
```

```
In [34]: type(mylist)
```

```
Out[34]: list
```

```
In [35]: intX = 1
```

```
In [36]: type(intX)
```

```
Out[36]: int
```

```
In [38]: BooleanX = True
```

```
In [39]: type(BooleanX)
```

```
Out[39]: bool
```



```
i = 10 # 設定控制變數
while i > 0:
    # 迴圈工作區
    print(i)
    i -= 1 # 調整控制變數值
```

```
for i in range(10, 0, -1):
    print(i)
```

```
In [50]: x = 1  
for y in range(1,10):  
    print (x, '*', y, '=', x*y)
```

```
1 * 1 = 1  
1 * 2 = 2  
1 * 3 = 3  
1 * 4 = 4  
1 * 5 = 5  
1 * 6 = 6  
1 * 7 = 7  
1 * 8 = 8  
1 * 9 = 9
```

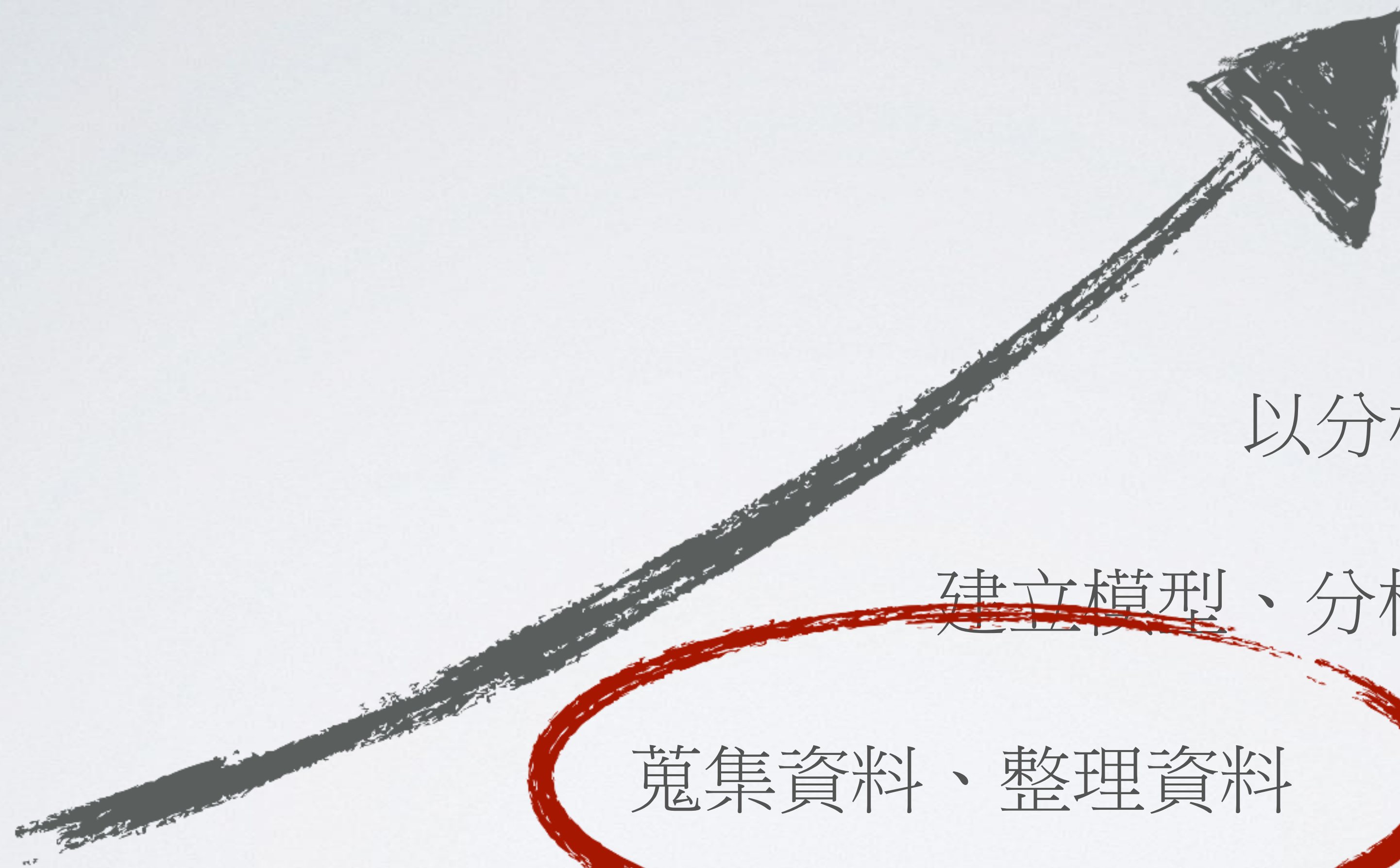

MARCH

S	M	T	W	T	F	S
28	29	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

```
In [50]: x = 1
for y in range(1,10):
    print (x, '*', y, '=', x*y)
```

```
1 * 1 = 1
1 * 2 = 2
1 * 3 = 3
1 * 4 = 4
1 * 5 = 5
1 * 6 = 6
1 * 7 = 7
1 * 8 = 8
1 * 9 = 9
```

```
>>>
1*1=1  2*1=2  3*1=3  4*1=4  5*1=5  6*1=6  7*1=7  8*1=8  9*1=9
1*2=2  2*2=4  3*2=6  4*2=8  5*2=10 6*2=12 7*2=14 8*2=16 9*2=18
1*3=3  2*3=6  3*3=9  4*3=12 5*3=15 6*3=18 7*3=21 8*3=24 9*3=27
1*4=4  2*4=8  3*4=12 4*4=16 5*4=20 6*4=24 7*4=28 8*4=32 9*4=36
1*5=5  2*5=10 3*5=15 4*5=20 5*5=25 6*5=30 7*5=35 8*5=40 9*5=45
1*6=6  2*6=12 3*6=18 4*6=24 5*6=30 6*6=36 7*6=42 8*6=48 9*6=54
1*7=7  2*7=14 3*7=21 4*7=28 5*7=35 6*7=42 7*7=49 8*7=56 9*7=63
1*8=8  2*8=16 3*8=24 4*8=32 5*8=40 6*8=48 7*8=56 8*8=64 9*8=72
1*9=9  2*9=18 3*9=27 4*9=36 5*9=45 6*9=54 7*9=63 8*9=72 9*9=81
>>> ===== RESTART =====
```

以分析結果回答問題

建立模型、分析資料

蒐集資料、整理資料

定義問題、規劃監圖

80%



[Browse](#)[Shows](#)[Reviews](#)[PS4](#)[Xbox](#)[PC](#)[Nintendo](#)[Movies](#)[TV](#)[Tech](#)[Esports](#)[Sign in](#) [Boards](#)[Gaming](#)[PC](#)[Jump to...](#)

PC

[Tweet](#)[G+](#)[Like](#)

15 people like this. Be the first of your friends.

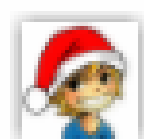
Page 1 of 1753 [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [→](#) [1753](#) [Next >](#)[Add to Favorites](#)[Log in](#)

TITLE

START DATE

REPLIES VIEWS

LAST MESSAGE ↓

[The PC board Steam/Origin & other online IDs thread](#)

Syfadious, MAR 21, 2013



Replies:

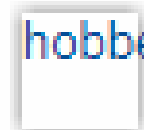
228

Views:

81,259

Magliris

Sep 14, 2017

[Community notice: Advertising](#)

hobbes259, MAY 1, 2013



Replies:

69

Views:

27,248

Cllzquierdo

Jun 21, 2017

[Game Discussion Index](#)

Perfect_Hand, OCT 19, 2015



Replies:

32

Views:

9,507

Benthe13th

May 22, 2017

[The IGN PC hardware guide and discussion thread](#)

Syfadious, APR 4, 2013



Replies:

630

Views:

133,367

FinalDrive

May 13, 2017

[GPU fail after install of drivers?](#)

Hydn1, YESTERDAY AT 8:22 AM

Replies:

2

Views:

107

Hydn1

Today at 3:08 AM

[Blu-ray player software???](#)

mfletch512, MAY 22, 2013

Replies:

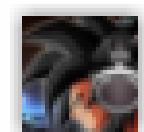
39

Views:

3,785

Player177

Today at 2:34 AM

[A recommended 1440p \(gsync\)](#)

jarjar111, YESTERDAY AT 11:36 AM

Replies:

2

Views:

93

Player177

Today at 2:29 AM

[Virus Gaming \(YouTube Gaming Channel\)](#)

virusgaming, WEDNESDAY AT 9:36 PM

Replies:

1

Views:

98

virusgaming

Yesterday at 7:53 PM

[What is up with PC gaming and no physical copies?](#)

Replies:

320

Views:

8,741

Closingracer

Yesterday at 4:15 PM

全國電子

冷氣震撼出清

再享15期0利率

(大金、日立、Panasonic、富士通品牌及限時限量特價商品除外)

點我看詳情

MEMBERS ONLINE NOW

Total: 0 (members: 0, guests: 0, robots: 0)

FORUM STATISTICS

Discussions: 0

Messages: 0

Members: 0

Latest Member:

[SHARE THIS PAGE](#)

	:::: The PC board Steam/Origin & other online IDs thread ::: Syfadious, MAR 21, 2013		Replies: 228 Views: 81,259	Magli Sep 14, 20
	Community notice: Advertising hobbes259, MAY 1, 2013		Replies: 69 Views: 27,248	Clizquiero Jun 21, 20
	Game Discussion Index Perfect_Hand, OCT 19, 2015		Replies: 32 Views: 9,507	Benthe13 May 22, 20
	:::: The IGN PC hardware guide and discussion thread ::: Syfadious, APR 4, 2013		Replies: 630 Views: 133,367	FinalDri May 13, 20

- Title
- Poster
- Time(Post time / Reply time)
- URL(To get all the post)

Browse

Search

ShowsReviewsPS4XboxPCNintendoMoviesTVTechEsports

Sign in

1 of 1753

1

2

3

4

5

6

→ 1753Next >

Jump to...

Tweet

G+

Like

15 people like this. Be the first of your friends.

TITLE	START DATE	REPLIES	VIEWS	LAST MESSAGE
...: The PC board Steam/Origin & other online IDs thread ...: Syfadious, MAR 21, 2013		Replies: Views: 81,259	228	Magliris Sep 14, 2017
... Community notice: Advertising hobbes259, MAY 1, 2013		Replies: Views: 27,248	69	Clizquierdo Jun 21, 2017
Game Discussion Index Perfect_Hand, OCT 19, 2015		Replies: Views: 9,507	32	Benthe13th May 22, 2017
...: The IGN PC hardware guide and discussion thread ...: Syfadious, APR 4, 2013		Replies: Views: 133,367	630	FinalDrive May 13, 2017
... thread-453045707.discussionListItem.visible 812 x 48		Replies: Views: 107	2	Hydn1 Today at 3:08 AM
Blu-ray player software??? mlfetch512, MAY 22, 2013		Replies: Views: 3,785	39	Player177 Today at 2:34 AM
A recommended 1440p (gsync) jarjar111, YESTERDAY AT 11:36 AM		Replies: Views: 93	2	Player177 Today at 2:29 AM
Virus Gaming (YouTube Gaming Channel) virusgaming, WEDNESDAY AT 9:36 PM		Replies: Views: 98	1	virusgaming Yesterday at 7:53 PM
What is up with PC gaming and no physical copies? Tony72495, SEP 23, 2015		Replies: Views: 8,741	320	Closingracer Yesterday at 4:15 PM
GTA V with mod ULTRA REALISTIC GRAPHICS cacaocaca1, YESTERDAY AT 4:02 PM		Replies: Views: 48	0	cacaocaca1 Yesterday at 4:02 PM
If you had to describe this outfit Centrinario, WEDNESDAY AT 7:37 AM		Replies: Views: 125	1	Centrinario Yesterday at 3:52 PM
Minecraft "Better together" (1.2.0) broken? Mandoble, SUNDAY AT 2:58 AM		Replies: Views: 430	8	Mandoble Yesterday at 3:13 PM
[PUBG] So I destroyed my door, the rage is getting me in game! Hollandkerel, YESTERDAY AT 2:32 PM		Replies: Views: 67	0	Hollandkerel Yesterday at 2:32 PM
Steam wallet codes Mr._A_Person, WEDNESDAY AT 1:24 PM		Replies: Views: 204	6	Mr._A_Person Yesterday at 1:41 PM
Horror Games DeepestSeas, MONDAY AT 4:25 PM		Replies: Views: 204	4	Employee427 Yesterday at 10:43 AM
The PC Games Discussion Thread siixpool, AUG 28, 2015		Replies: Views: 94,227	1,970	Employee427 Yesterday at 10:36 AM
New to Pc looking for steam friends MordxFustang, SUNDAY AT 8:44 PM		Replies: Views: 254	7	jarjar111 Yesterday at 10:33 AM

Log in

MEMBERS ONLINE NOW

Total: 0 (members: 0, guests: 0, robots: 0)

FORUM STATISTICS

Discussions: 0

Messages: 0

Members: 0

Latest Member:

SHARE THIS PAGE

Tweet

Like 298

G+

ElementsConsoleSourcesNetworkPerformanceMemoryApplicationSecurityAudits

<div class="mainContainer">

<div class="mainContent">

<!-- main template -->

<form id="favForumForm" action="account/favorite-forum/add" method="post" class="xenForm AutoValidator">...</form>

<div class="pageNavLinkGroup">...</div>

<div class="discussionList section sectionMain">

<form action="inline-mod/thread/switch" method="post" class="DiscussionList InlineModForm" data-cookieName="threads" data-controls="#InlineModControls" data-imodoptions="#ModerationSelect option">

<dl class="sectionHeaders">...</dl>

<ol class="discussionListItems"> == \$0

<li id="thread-452928394" class="discussionListItem visible sticky" data-author="Syfadious">...

<li id="thread-452992087" class="discussionListItem visible sticky" data-author="hobbes259">...

<li id="thread-454612693" class="discussionListItem visible sticky" data-author="Perfect_Hand">...

<li id="thread-452947298" class="discussionListItem visible sticky" data-author="Syfadious">...

<li class="sectionHeaders">

<li id="thread-455032951" class="discussionListItem visible" data-author="Hydn1">...

<li id="thread-453045707" class="discussionListItem visible" data-author="mlfetch512">...

<li id="thread-455033022" class="discussionListItem visible" data-author="jarjar111">...

<li id="thread-455032839" class="discussionListItem visible" data-author="virusgaming">...

<li id="thread-454591558" class="discussionListItem visible" data-author="Tony72495">...

<li id="thread-455033125" class="discussionListItem visible" data-author="cacaocaca1">...

<li id="thread-455032514" class="discussionListItem visible" data-author="Centrinario">...

<li id="thread-455031202" class="discussionListItem visible" data-author="Mandoble">...

<li id="thread-455033081" class="discussionListItem visible" data-author="Hollandkerel">...

<li id="thread-455032625" class="discussionListItem visible" data-author="Mr._A_Person">...

<li id="thread-455031855" class="discussionListItem visible" data-author="DeepestSeas">...

<li id="thread-454568685" class="discussionListItem visible" data-author="siixpool">...

<li id="thread-455031564" class="discussionListItem visible" data-author="MordxFustang">...

<li id="thread-455032531" class="discussionListItem visible" data-author="Bird_Thing">...

<li id="thread-455031308" class="discussionListItem visible" data-author="thicket04">...

<li id="thread-455032493" class="discussionListItem visible" data-author="Qelta">...

<li id="thread-455030116" class="discussionListItem visible" data-author="brian.vespucci">...

<li id="thread-455031199" class="discussionListItem visible" data-author="SaltSlasher">...

<li id="thread-455028235" class="discussionListItem visible" data-author="frank0393">...

<li id="thread-455018112" class="discussionListItem visible" data-author="Pana-Wan_Kenobi">...

<li id="thread-455026782" class="discussionListItem visible" data-author="Boice">...

<li id="thread-455027959" class="discussionListItem visible" data-author="ViipartyGames">...

<li id="thread-454790165" class="discussionListItem visible" data-author="Pana-Wan_Kenobi">...

StylesComputed

Filter: :hov .cls

element.style { }

css.php?css=dis...150665244... .discussionListItems { }

css.php?css=xen...d=15066524... body, div, dl, dt, dd, ul, ol, li, h1, h2, h3, h4, h5, h6, pre, code, form, fieldset, legend, input, button, textarea, p, blockquote, th, td { margin: 0; padding: 0; }

ol { user agent stylesheet display: block; list-style-type: decimal; -webkit-margin-before: 1em; -webkit-margin-after: 1em; -webkit-margin-start: 0px; -webkit-margin-end: 0px; -webkit-padding-start: 40px; }

Inherited from body.node7203... css.php?css=xen...150665244... html body { font-family: sans-serif; min-width: initial; }

css.php?css=xen...d=15066524... body { -webkit-text-size-adjust: 100%; -moz-text-size-adjust: 100%; -ms-text-size-adjust: 100%; text-size-adjust: 100%; font-family: "Helvetica Neue", Helvetica, Arial, serif; color: rgb(20,20,20); word-wrap: break-word; line-height: 1.27; }

css.php?css=xen...d=15066524... body { font: 13px/1.231 arial,helvetica, clean, serif; }

Inherited from html#XenForo... css.php?css=xen...d=15066524... html { }

... ol.discussionListItems li#thread-455032951.discussionListItem.visible div.listBlock.main div h3 a

ConsoleWhat's New

Highlights from the Chrome 61 update

檢視原始碼

選定目標

重複性目標

http://www.ign.com/boards/forums/pc.7203/

IGN Boards PC forum page showing a list of threads. The threads are sorted by replies. The first thread is "The PC board Steam/Origin & other online IDs thread" by Syfadios, dated MAR 21, 2013, with 228 replies and 81,259 views. The last thread is "Horror Games" by DeepestSeas, dated MONDAY AT 4:25 PM, with 4 replies and 204 views.

TITLE	START DATE	REPLIES	VIEWS	LAST MESSAGE
... The PC board Steam/Origin & other online IDs thread ... Syfadios, MAR 21, 2013		Replies: 228 Views: 81,259	Magliris Sep 14, 2017	
Community notice: Advertising hobbes259, MAY 1, 2013		Replies: 69 Views: 27,248	Clizquierdo Jun 21, 2017	
Game Discussion Index Perfect_Hand, OCT 19, 2015		Replies: 32 Views: 9,507	Benthe13th May 22, 2017	
... The IGN PC hardware guide and discussion thread ... Syfadios, APR 4, 2013		Replies: 630 Views: 133,367	FinalDrive May 13, 2017	
li#thread-453045707 discussionListItem visible 812x48		Replies: 2 Views: 107	Hydn1 Today at 3:08 AM	
Blu-ray player software??? mfletch512, MAY 22, 2013		Replies: 39 Views: 3,785	Player177 Today at 2:34 AM	
A recommended 1440p (gsync) jarjar111, YESTERDAY AT 11:36 AM		Replies: 2 Views: 93	Player177 Today at 2:29 AM	
Virus Gaming (YouTube Gaming Channel) virusgaming, WEDNESDAY AT 7:53 PM		Replies: 1 Views: 98	virusgaming Yesterday at 7:53 PM	
What is up with PC gaming and no physical copies? Tony72495, SEP 23, 2015		Replies: 320 Views: 8,741	Closingracer Yesterday at 4:15 PM	
GTA V with mod ULTRA REALISTIC GRAPHICS cacaocaca1, YESTERDAY AT 4:02 PM		Replies: 0 Views: 48	cacaocaca1 Yesterday at 4:02 PM	
If you had to describe this outfit Centrinario, WEDNESDAY AT 7:37 AM		Replies: 1 Views: 125	Centrinario Yesterday at 3:52 PM	
Minecraft "Better together" (1.2.0) broken? Mandoble, SUNDAY AT 2:58 AM		Replies: 8 Views: 430	Mandoble Yesterday at 3:13 PM	
[PUBG] So I destroyed my door, the rage is getting me in game! Hollandkerel, YESTERDAY AT 2:32 PM		Replies: 0 Views: 67	Hollandkerel Yesterday at 2:32 PM	
Steam wallet codes Mr._A_Person, WEDNESDAY AT 1:24 PM		Replies: 6 Views: 204	Mr._A_Person Yesterday at 1:41 PM	
Horror Games DeepestSeas, MONDAY AT 4:25 PM		Replies: 4 Views: 204	Employee427 Yesterday at 10:43 AM	

Browser developer tools showing the HTML source code of the forum page. The code is structured as follows:

```
<div class="mainContent">
  <!-- main template -->
  <form id="favForumForm" action="account/favorite-forum/add" method="post"
  class="xenForm AutoValidator">...</form>
  <div class="pageNavLinkGroup">...</div>
  <div class="discussionList section sectionMain">
    <form action="inline-mod/thread/switch" method="post" class="DiscussionList
    InlineModForm" data-cookieName="threads" data-controls="#InlineModControls"
    data-imodOptions="#ModerationSelect option">
      <dl class="sectionHeaders">...</dl>
      <div class="discussionListItems">
        <li id="thread-452928394" class="discussionListItem visible sticky"
        data-author="Syfadios">...</li>
        <li id="thread-452992087" class="discussionListItem visible sticky"
        data-author="hobbes259">...</li>
        <li id="thread-454612693" class="discussionListItem visible sticky"
        data-author="Perfect_Hand">...</li>
        <li id="thread-452947298" class="discussionListItem visible sticky"
        data-author="Syfadios">...</li>
        <li class="sectionHeaders">...</li>
        <li id="thread-455032951" class="discussionListItem visible" data-
        author="Hydn1">...</li>
        <li id="thread-453045707" class="discussionListItem visible" data-
        author="mfletch512">...</li>
        <li id="thread-455033022" class="discussionListItem visible" data-
        author="jarjar111">...</li>
        <li id="thread-455032839" class="discussionListItem visible" data-
        author="virusgaming">...</li>
        <li id="thread-454591558" class="discussionListItem visible" data-
        author="Tony72495">...</li>
        <li id="thread-455033125" class="discussionListItem visible" data-
        author="cacaocaca1">...</li>
        <li id="thread-455032514" class="discussionListItem visible" data-
        author="Centrinario">...</li>
        <li id="thread-455031202" class="discussionListItem visible" data-
        author="Mandoble">...</li>
        <li id="thread-455033081" class="discussionListItem visible" data-
        author="Hollandkerel">...</li>
        <li id="thread-455032625" class="discussionListItem visible" data-
        author="Mr._A_Person">...</li>
        <li id="thread-455031855" class="discussionListItem visible" data-
        author="DeepestSeas">...</li>
        <li id="thread-454568685" class="discussionListItem visible" data-
        author="siixpool">...</li>
        <li id="thread-455031564" class="discussionListItem visible" data-
        author="MordxFustang">...</li>
        <li id="thread-455032531" class="discussionListItem visible" data-
        author="Bird_Thing">...</li>
        <li id="thread-455031308" class="discussionListItem visible" data-
        author="thicket04">...</li>
        <li id="thread-455032493" class="discussionListItem visible" data-
        author="Qelta">...</li>
        <li id="thread-455030116" class="discussionListItem visible" data-
        author="brian.vespucci">...</li>
        <li id="thread-455031199" class="discussionListItem visible" data-
        author="SaltSlasher">...</li>
        <li id="thread-455028235" class="discussionListItem visible" data-
        author="frank0393">...</li>
        <li id="thread-455018112" class="discussionListItem visible" data-
        author="Pana-Wan_Kenobi">...</li>
        <li id="thread-455026782" class="discussionListItem visible" data-
        author="Boice">...</li>
        <li id="thread-455027959" class="discussionListItem visible" data-
```

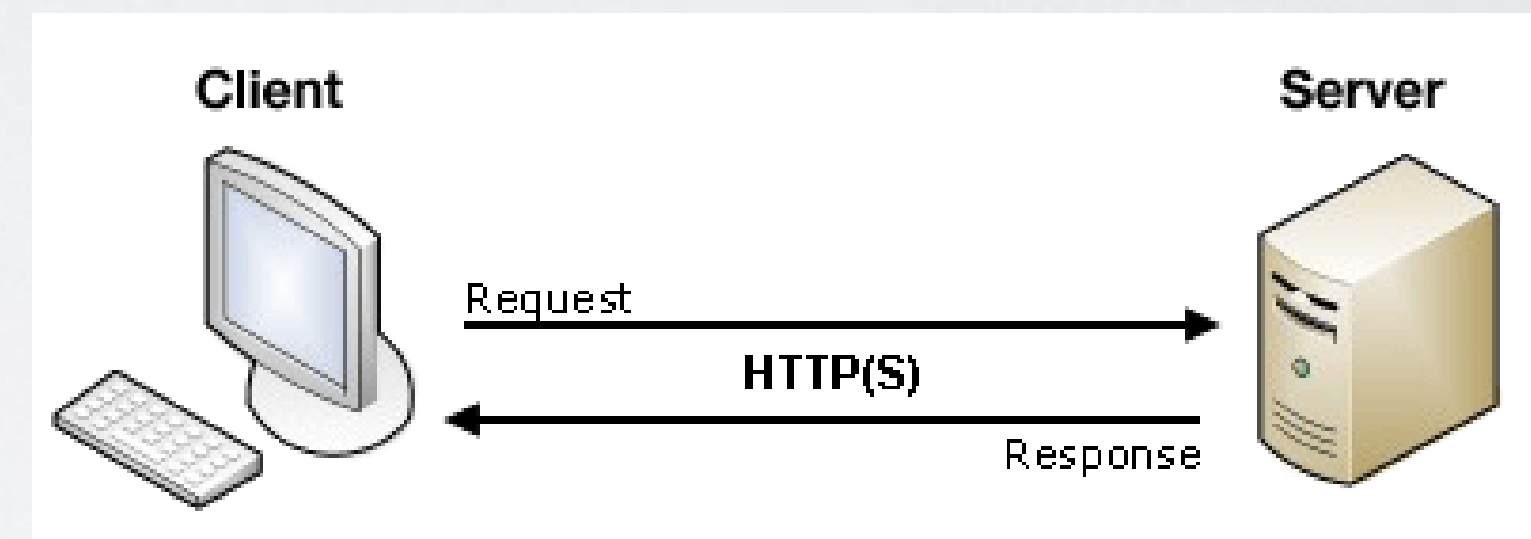


```
In [2]: import requests
```

```
In [3]: from bs4 import BeautifulSoup
```

```
In [4]: import pandas as pd
```

功能：模擬 html request




```
In [2]: import requests
```

```
In [3]: from bs4 import BeautifulSoup
```

```
In [4]: import pandas as pd
```

功能：
解析處理導航、
搜索、修改分析樹等功能。

```
In [2]: import requests
```

```
In [3]: from bs4 import BeautifulSoup
```

```
In [4]: import pandas as pd
```

功能：

Pandas 主要提供快速便捷地處理
結構化數據 的 大量數據結構 和 函數

你看到了啥??

```
In [14]: url = 'http://www.ign.com/boards/forums/pc.7203/'  
         response = requests.get(url)  
         print(response.text)
```

看到這些 Tag 標籤

```
In [14]: url = 'http://www.ign.com/boards/forums/pc.7203/'  
response = requests.get(url)  
print(response.text)
```

```
</div>  
<a class="queue-item promoted hidden" id="promoted-spine-dfp-container" href=""><span class="hidde  
n">Advertisement</span>  
<div id="pawn33" class="chess"></div></a>
```

```
<a class="queue-item trending " href="/articles/2017/09/29/snes-classic-where-to-buy">  
  <div class="queue-item-content">  
    <div class="title">SNES Classic: Where to Buy</div>  
  </div>  
</a>
```

```
<a class="queue-item trending " href="/articles/2017/09/29/blade-runner-2049-review">
```


取得網頁







```
In [ ]: url = 'http://www.ign.com/boards/forums/pc.7203/'
        response = requests.get(url)
```

```
In [ ]: #呼叫 BeautifulSoup
        soup = BeautifulSoup(response.text, 'lxml')
```

lxml 用來解析 html/xml

```
In [ ]: #截取整個"討論條"
        articles2 = soup.find_all(attrs={'class':'discussionListItem'})
        # attrs={'class':'discussionListItem visible'}
        # 檢查一下抓到幾筆
        print(len(articles2))
```

soup.find_all
soup.find

	... The IGN PC hardware guide and discussion thread ... Svfadious APR 4, 2013		Replies: 630 Views: 133,367	FinalDrive May 13, 2017
	GPU fail after install of drivers? Hydn1, YESTERDAY AT 8:22 AM		Replies: 2 Views: 107	Hydn1 Today at 3:08 AM
	Blu-ray player software???		Replies: 39 Views: 3,785	Player177 Today at 2:34 AM
	A recommended 1440p (gsync) jarjar111, YESTERDAY AT 11:36 AM		Replies: 2 Views: 93	Player177 Today at 2:29 AM
	Virus Gaming (YouTube Gaming Channel) virusgaming, WEDNESDAY AT 9:36 PM		Replies: 1 Views: 98	virusgaming Yesterday at 7:53 PM

再享15期0利率
(大金、日立、Panasonic、富士通品牌及限時限量特價商品除外)
點我看詳情

MEMBERS ONLINE NOW
Total: 0 (members: 0, guests: 0, robots: 0)

FORUM STATISTICS

Discussions:	0
Messages:	0
Members:	0
Latest Member:	

```
<li id="thread-455032951" class="discussionListItem visible" data-
author="Hydn1"> == $0
  <div class="listBlock posterAvatar">...</div>
  <div class="listBlock main">
    <div class="titleText">
      <h3 class="title">
        <a href="threads/gpu-fail-after-install-of-drivers.455032951/"
          title class="PreviewTooltip" data-previewurl="threads/gpu-fail-
          after-install-of-drivers.455032951/preview">GPU fail after
          install of drivers?</a>
      </h3>
      <div class="secondRow">...</div>
    </div>
    <div class="listBlock stats pairsJustified" title="Members who liked
the first message: 0">...</div>
    <div class="listBlock lastPost">...</div>
  </div>
```

```
<li id="thread-455032951" class="discussionListItem visible"
author="Hydn1"> == $0
```

```
<div class="listBlock posterAvatar">...</div>
```

```
<div class="listBlock main">
```

```
<div class="titleText">
```

取得網頁

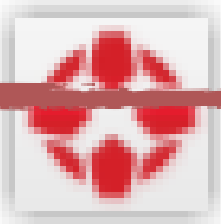
```
] : url = 'http://www.ign.com/boards/forums/pc.7203/'  
    response = requests.get(url)
```

```
] : # 呼叫 BeautifulSoup  
    soup = BeautifulSoup(response.text, 'lxml')
```

```
] : # 截取整個"討論條"  
    articles2 = soup.find_all(attrs={'class': 'discussionListItem'})  
    # attrs={'class': 'discussionListItem visible'}  
    # 檢查一下抓到幾筆  
    print(len(articles2))
```


Title

Url



GPU fail after install of drivers?

Hydn1

Replies: 2

Today at 3:08 AM

```
<li class="sectionHeaders">&nbsp;</li>
▼<li id="thread-455032951" class="discussionListItem visible " data-author="Hydn1">
  ▶<div class="listBlock posterAvatar">...</div>
  ▼<div class="listBlock main">
    ▼<div class="titleText">
      ▼<h3 class="title">
        <a href="threads/gpu-fail-after-install-of-drivers.455032951/" title class="
        "PreviewTooltip" data-previewurl="threads/gpu-fail-after-install-of-drivers.455032951/
        preview">GPU fail after install of drivers?</a> == $0
      </h3>
      ▼<div class="secondRow">
        ▼<div class="posterDate muted">
          <a href="members/hydn1.96496889/" class="username" dir="auto" title="Thread starter">
            Hydn1</a>
          ▶<span class="startDate">...</span>
        </div>
      </div>
    </div>
  </li>
```

```
: # 在articles2 建立完後 接著找出所有要的資料

for articles2 in articles2:
#     擷取所有Title
    title2 = articles2.find('a', 'PreviewTooltip').getText().strip()
#     擷取超連結
    link2 = articles2.find('a', 'PreviewTooltip').get('href')
```



```
In [ ]: # 在articles2 建立完後 接著找出所有要的資料

for articles2 in articles2:
#     擷取所有Title
    title2 = articles2.find('a', 'PreviewTooltip').getText().strip()
#     擷取超連結
    link2 = articles2.find('a', 'PreviewTooltip').get('href')
```

```
<li class="sectionHeaders">&nbsp;</li>
▼<li id="thread-455032951" class="discussionListItem visible " data-author="Hydn1">
▶<div class="listBlock posterAvatar">...</div>
▼<div class="listBlock main">
    ▼<div class="titleText">
        ▼<h3 class="title">
            <a href="threads/gpu-fail-after-install-of-drivers.455032951/" title class=
            "PreviewTooltip" data-previewurl="threads/gpu-fail-after-install-of-drivers.455032951/
            preview">GPU fail after install of drivers?</a> == $0
            </h3>
        ▼<div class="secondRow">
            ▼<div class="posterDate muted">
                <a href="members/hydn1.96496889/" class="username" dir="auto" title="Thread starter">
                Hydn1</a>
                ▶<span class="startDate">...</span>
```



::: The IGN PC hardware guide and discussion thread :::

Svfadious, APR 4, 2013



Replies:

630

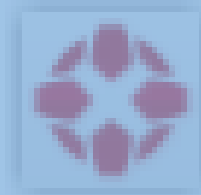
Views:

133,367

FinalDrive

May 13, 2017

li#thread-455032951.discussionListItem.visible | 812 x 48



GPU fail after install of drivers?

Hydn1, YESTERDAY AT 8:22 AM

Replies:

2

Views:

107

Hydn1

Today at 3:08 AM



Blu-ray player software???

mfetch512, MAY 22, 2013

Replies:

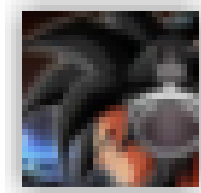
39

Views:

3,785

Player177

Today at 2:34 AM



A recommended 1440p (gsync)

jarjar111, YESTERDAY AT 11:36 AM

Replies:

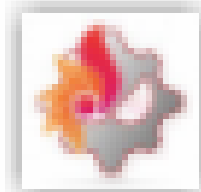
2

Views:

93

Player177

Today at 2:29 AM



Virus Gaming (YouTube Gaming Channel)

virusgaming, WEDNESDAY AT 9:36 PM

Replies:

1

Views:

53

virusgaming

Yesterday at 7:53 PM

```
<a class="faint">
```

```
<span class="DateTime" title>May 22, 2013</span>
```

```
</a>
```

```
</span>
```

```
<a class="dateTime">
```

```
<abbr class="DateTime" data-time="1506677662" data-  
diff="6375" data-datestring="Sep 29, 2017" data-  
timestring="2:34 AM" title="Sep 29, 2017 at 2:34  
AM">Today at 2:34 AM</abbr> == $0
```

```
</a>
```



```
In [ ]: # 在articles2 建立完後 接著找出所有要的資料

for articles2 in articles2:
    # 擷取所有Title
    title2 = articles2.find('a', 'PreviewTooltip').getText().strip()
    # 擷取超連結
    link2 = articles2.find('a', 'PreviewTooltip').get('href')
    # 截取使用發文時間
    mainItemTime = articles2.find('abbr', 'DateTime')
    # 另外處理最後發文時間，它在另一個區塊裡
    lastPos2 = articles2.find('dl', 'lastPostInfo')
    # 判斷時間，因為屬性不一樣
    if articles2.find('abbr', 'DateTime')== None:
        dateTime2 = articles2.find('span', 'DateTime').getText()
        lastPosTime = lastPos2.find('span', 'DateTime').getText()
    else:
        dateTime2 = articles2.find('abbr', 'DateTime').getText()
        lastPosTime = lastPos2.find('abbr', 'DateTime').getText()
    print(title2) # result of setp-3
    print("http://www.ign.com/boards/"+link2)
    print(dateTime2,lastPosTime)
    print("====")
```

所有文章

#1

Advertisement

Ads by Google ⓘ




isawntem likes this

```
▼<blockquote class="messageText SelectQuoteContainer ugc  
baseHtml"> == $0
```



Soooo I just got a blu-ray player for my computer but for some reason that is beyond me, it didnt actually come with the software to play Blu-ray movies... "



so, now i have this drive and it is basically useless."



can you guys help me find a software that will actually play blu-rays for me, preferably one that is free if possible. I read about GOM Player but it wouldnt open the



Browse

Search

Shows Reviews PS4 Xbox PC Nintendo Movies TV Tech Esports

Sign in

Discussion in 'PC' started by mfletch512, May 22, 2013.



mfletch512
Almost Not a Noob

Joined: Jul 18, 2012
Messages: 737
Location: Michigan

Date Posted: May 22, 2013

#1

Soooo I just got a blu-ray player for my computer but for some reason that is beyond me, it didnt actually come with the software to play Blu-ray movies...
so, now i have this drive and it is basically useless.
can you guys help me find a software that will actually play blu-rays for me, preferably one that is free if possible. I read about GOM Player but it wouldnt open the file and I couldnt find the right Codecs. if you know where i can get the Codecs for them that would be great too.

Advertisement

Ad closed by Google

Stop seeing this ad

Ads by Google

mfletch512, May 22, 2013

isawntFM likes this

[]:

擷取内容function

def getMessageContent(url):

messageContentList=[]

response = requests.get(url)

soup = BeautifulSoup(response.text, 'lxml')

blockquote = soup.find_all('blockquote', 'messageText SelectQuoteContainer ugc baseHtml')

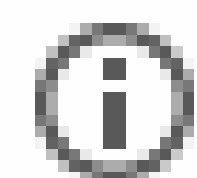
len(articles)

for blockquote in blockquote:

message = blockquote.getText().strip()

messageContentList.append(message)

return messageContentList



www.ign.com/boards/threads/blu-ray-player-software.453045707/

▼<article>

▼<blockquote class="messageText SelectQuoteContainer ugc baseHtml"> == \$0

"

Soooo I just got a blu-ray player for my computer but for some reason that is beyond me, it didnt actually come with the software to play Blu-ray movies... "

"

Data Frame ? why

```
In [ ]: # 建立資料表
df = pd.DataFrame(allarticle, columns=["title2", "link2", "onePost", "dateTime2", "lastPosTime"])
# 輸出CSV
df.to_csv('list.csv', index=False)
```

```
In [40]: pttArticleTitle
```

```
Out[40]: [['[好雷] 《老師你會不會回來》，我們不會被地震', '9/29', 'a122239', '9'],
          ['[討論] 希望哪款手機遊戲改編電影', '9/29', 'zy116pj', '21'],
          ['[妙雷] 《母親》太怪了啦，不曉得該說什麼', '9/29', 'eytdtdm8', '10'],
          ['[普好雷] 《金牌特務：機密對決》-便當店喬遷', '9/29', 'kielkiel', '5'],
          ['[好雷] 真正的騎士精神！金牌特務3的必點！', '9/29', 'Tang-sing', '11']]
```

```
In [41]: df
```

```
Out[41]:
```

	title	date	author	push
0	[好雷] 《老師你會不會回來》，我們不會被地震	9/29	a122239	8
1	[討論] 希望哪款手機遊戲改編電影	9/29	zy116pj	21
2	[妙雷] 《母親》太怪了啦，不曉得該說什麼	9/29	eytdtdm8	10
3	[普好雷] 《金牌特務：機密對決》-便當店喬遷	9/29	kielkiel	5

實戰 2

PTT

批踢踢實業坊 > 看板 movie

聯絡資訊 關於我們

看板

精華區

最舊

< 上頁

下頁 >

最新

- 8 [\[好雷\] 《老師你會不會回來》，我們不會被地震](#)
9/29 a122239
- 21 [\[討論\] 希望哪款手機遊戲改編電影](#)
9/29 zy116pj
- 8 [\[妙雷\] 《母親》太怪了啦，不曉得該說什麼](#)
9/29 eytdtdm8
- 5 [\[普好雷\] 《金牌特務：機密對決》-便當店喬遷](#)
9/29 kielkiel
- [\[好雷\] 真正的騎士精神?金牌特務2的幽默](#)
9/29 Tenacious
- [\[問片\] 15夜之歌](#)
9/29 ha602109
- 2 [\[新聞\] 行銷西拉雅 來看全台1/4老鷹集中地](#)
9/29 iam168888888
- 1 [\[新聞\] 金馬54年度廣告 杜琪\(峰\)操刀想像2117年](#)
9/29 iam168888888
- [\[LIVE\] CINEMAX 22:00 亡命賽車2050 \(Death Race](#)
9/29 pop0311
- 3 [\[負雷\] 金牌特務2還我大型煙火秀](#)
9/29 iam168888888

最新

X 22:00 亡命賽車2050 (Death Race

► <div class="r-ent">_</div>

```
In [19]: # 宣告
import requests
from bs4 import BeautifulSoup

# 所有東西放這裡
pttArticleTitle = []

# 鎖定網址
url = 'https://www.ptt.cc/bbs/movie/index5885.html'
print(url)
response = requests.get(url)
# 宣告湯~~~~~
soup = BeautifulSoup(response.text, 'lxml')
# 找尋文章區塊
articles = soup.find_all('div', 'r-ent')

# 進行爬蟲
for article in articles:
# 區塊
    meta = article.find('div', 'title').find('a')
# 標題
    title = meta.getText().strip()
# Url
    link = meta.get('href')
# 推文數量
    push = article.find('div', 'nrec').getText()
# 發文時間
    date = article.find('div', 'date').getText()
# 作者
    author = article.find('div', 'author').getText()
    print(title,date,author,push) # result of setp-3
    title=[title,date,author,push]
    pttArticleTitle.append(title)
```



```
# URL編號放這裡
urls = ['5885', '5884', '5883', '5881', '5880', '5879', '5878', '5877', '5876', '5875']
# 所有東西放這裡
pttArticleTitle = []
```

原本的

```
# 鎖定網址
url = 'https://www.ptt.cc/bbs/movie/index5885.html'
print(url)
response = requests.get(url)
# 宣告湯~~~~~
soup = BeautifulSoup(response.text, 'lxml')
# 找尋文章區塊
articles = soup.find_all('div', 'r-ent')

# 進行爬蟲
for article in articles:
    # 區塊
    meta = article.find('div', 'title').find('a')
    # 標題
    title = meta.getText().strip()
    # Url
    link = meta.get('href')
    # 推文數量
```

新加的

```
for url in urls:
    # 鎖定網址
    url = 'https://www.ptt.cc/bbs/movie/index'+url+'.html'
    print(url)
    response = requests.get(url)
    # 宣告湯~~~~~
    soup = BeautifulSoup(response.text, 'lxml')
    # 找尋文章區塊
    articles = soup.find_all('div', 'r-ent')
```

!!!!!!??????


```
▼ <div class="r-ent">
```

```
▶ <div class="nrec">...</div>
```

```
  <div class="mark"></div>
```

```
▼ <div class="title">
```

```
  <a href="/bbs/movie/M.1506589068.A.6AE.html">[新聞] 索尼經典購得Glenn Close的「The Wife」 </a>
```

```
  </div>
```

```
▶ <div class="meta">...</div>
```

```
</div>
```

```
▼ <div class="r-ent">
```

```
▶ <div class="nrec">...</div>
```

```
  <div class="mark"></div>
```

```
  <div class="title">
```

(本文已被刪除) [BlueTuna601]

```
  </div> == $0
```

```
▶ <div class="meta">...</div>
```

```
# print(title, date, author, push) # result of step-3
title=[title,date,author,push]
pttArticleTitle.append(title)
```

```
https://www.ptt.cc/bbs/movie/index5885.html
https://www.ptt.cc/bbs/movie/index5884.html
https://www.ptt.cc/bbs/movie/index5883.html
https://www.ptt.cc/bbs/movie/index5881.html
```

```
-----
AttributeError                                Traceback (most recent call last)
<ipython-input-22-629d7cb80b01> in <module>()
     25         meta = article.find('div', 'title').find('a') #or NOT_EXIST
     26         # 標題
--> 27         title = meta.getText().strip()
     28         # Url
     29         link = meta.get('href')
```

```
AttributeError: 'NoneType' object has no attribute 'getText'
```



```
NOT_EXIST = BeautifulSoup('<a>本文已被刪除</a>', 'lxml').a
```

```
for urls in urls:
    # 鎖定網址
    url = 'https://www.ptt.cc/bbs/movie/index'+urls+'.html'
    print(url)
    response = requests.get(url)
    # 宣告湯~~~~~
    soup = BeautifulSoup(response.text, 'lxml')
    # 找尋文章區塊
    articles = soup.find_all('div', 'r-ent')
    |
    # 進行爬蟲
    for article in articles:
        # 區塊
        meta = article.find('div', 'title').find('a') or NOT_EXIST
        # 標題
        title = meta.getText().strip()
        # Url
```

匯出

```
[21]: import pandas as pd
# 建立資料表
df = pd.DataFrame(pttArticleTitle, columns=["title", "date", "author", "push"])
# 輸出CSV
df.to_csv('pttlist.csv', index=False)
```

A	B	C	D
title	date	author	push
[好雷] 《老師你會不會回來》，我們不會被地震	9/29	a122239	8
[討論] 希望哪款手機遊戲改編電影	9/29	zy116pj	21
[妙雷] 《母親》太怪了啦，不曉得該說什麼	9/29	eytdtdm8	10
[普好雷] 《金牌特務：機密對決》-便當店喬遷	9/29	kielkiel	5
[好雷] 真正的騎士精神?金牌特務2的幽默	9/29	Tenacious	
[問片] 15夜之歌	9/29	ha602109	
[新聞] 行銷西拉雅 來看全台1/4老鷹集中地	9/29	iam168888	2
[新聞] 金馬54年度廣告 杜琪(峰)操刀想像2117年	9/29	iam168888	1
[LIVE] CINEMAX 22:00 亡命賽車2050 (Death Race	9/29	pop0311	
[好雷] 金牌特務2還我士型煙火季	9/29	ioci	3

