

數據支持真相，真相戳破謊言

「數據、謊言與真相」

專書閱讀心得

目 錄

一、前言	3
二、內容概要	
(一)、別讓直覺扯你的後腿	3
(二)、大數據的驚人力量	4
(三)、小心面對大數據	5
(四)、大數據告訴我，很少人看到最後一頁	6
三、啟發與心得	
(一)、一舉一動都逃不出大數據的天眼	6
(二)、數據翻轉謊言	7
(三)、讀數據的驚人成果-汽車貨運調查	9
四、結語:數據不是萬能，謊言卻是解藥	10
五、參考文獻	10

一、前言

什麼是大數據？書中並未對大數據下一個清楚的定義，在資訊爆炸時代，人類可能一天產生好幾萬兆的資訊量，要從數據中獲得寶貴資訊，最關鍵的一點是：你必須問對問題。而大數據有四大關鍵力量可以協助：1、大數據能讓你將數據切割分解，讓你見微知著及獲得具體的見解。2、新的資料來源通常包括新類型的變數，比現存變數更能幫助我們了解複雜關係，並充分利用。3、新的數位資訊能提供我們生活的真實樣貌，而非我們希望自己呈現給外人的形象。4、大數據易於與實驗結合，使我們能測試因果關係，而非僅是相關性。

進入網路新時代的這幾十年以來，我們在健康、道德、勞動市場、商業、恐怖主義、性別，以及種族等議題上有長足的認識。我們的數位足跡已把整個世界變成一個實驗室，這場數位革命如何為我們每個人打開找尋隱藏真理的大門，一窺人們的內心世界，提出大數據時代真正的洞見。

作者以極具啟發性的觀點解讀大數據，並以各式各樣有趣的案例忠實呈現網路世界與當代社會的現況，讓我們得以重新認識自己與這個世界。

二、內容概要

(一)、別讓直覺扯你的後腿

是什麼造就了NBA球星，在貧窮家庭長大，還是在中產階級家庭長大，讓你更有可能在NBA成名？大多數會猜前者。一般人都認為，在貧困環境下長大的，也許由十幾歲的未婚媽媽辛苦扶養的環境，有助於培養在這種競爭激烈運動比賽中達到頂尖水準所需的驅動力。費城高中籃球教練威

廉·艾勒比在接受《運動畫刊》採訪時表示：「郊區的小孩打球往往只是為了好玩，但都市貧民區的小孩卻把籃球當成生死攸關的大事。」，這種想法普遍正確嗎？利用許多來源的數據並檢測，可以弄清楚哪些家庭的背景最有利於 NBA 球員的養成，第一個相關數據是每個球員的出生地，接下來是出生在富裕郡的男孩，這表示事實跟普遍的想法正好相反，一但我們發現違反直覺的結果，我們更可以使用更多的數據科學來幫助我們解釋，為什麼這世界跟表面上看起來並不相同。

(二)、大數據的驚人力量

大數據的驚人力量：1、提供新類型的數據，就是大數據擁有的第一種力量。2、提供誠實的數據，就是大數據擁有的第二種力量。3、允許我們把焦點放在人口中的小子集，就是大數據擁有的第三種力量。4、允許我們進行許多因果關係的實驗，就是大數據擁有的第四種力量。對佛洛依德的分析，提供了很好的例證。

大數據擁有的第一種力量：重新想像什麼東西有資格當成數據。通常大數據的價值不在於本身的大小，而在於提供新類型的資訊，也就是以往從未蒐集過的資訊，讓人們得以研究。在某段時間裡有幾兆個 Google 搜尋，你認為跟失業最相關的搜尋是什麼？「就業服務中心」、「新工作」、「失業率」這些搜尋的頻率也很高，但不是最高，反而是色情網站或蜘蛛紙牌搜尋出現的最多，這並不代表色情網站或蜘蛛紙牌的搜尋次數等於失業率，

但是可以做為預測失業率最佳模式的一部分。

大數據擁有的第二種力量：某些線上來源讓人們承認在其他地方不願意承認的事情。這類數據的作用就像數位誠實豆沙包，以 Goole 搜尋為例，記得人們更誠實的條件嗎？是上網嗎？獨自一人？沒有人進行調查？都答對了。Goole 搜尋還有另一個巨大的優勢，讓人們說出真相那就是：誘因。你沒有誘因在調查時承認某些事，但你確實有誘因以 Goole 搜尋自己的在意的事情。

大數據擁有的第三種力量：大數據允許我們有意義地放大檢視數據集的細部，獲取關於我們是誰的新見解。除了年齡之外，我們還可以放大其他面向，如果有足夠的數據，可以觀察特定城鎮居民的行為，觀察人們每小時，甚至每分鐘都做了什麼。

大數據擁有的第四種力量：大數據讓隨機對照實驗，也就是可以找到真正因果關係的方法，變得更容易進行。相較於現實世界的離線實驗，數位世界的實驗具有龐大的優勢，只要你在線上，幾乎隨時隨地都可以進行。在大數據時代，整個世界就是一間實驗室。

(三)、小心面對大數據

新的大數據源可以成功預測股票走勢嗎？簡單就是：不能。大數據是有所侷限的，我們無法利用大數據做什麼，以及有時我們不應該拿大數據來做什麼。過分迷戀大數據將造成盲點，人們使用數據做決定時，會遇到

一個潛在的問題：數字可能誘惑我們。如果愈來愈迷戀數據，可能會讓我們忽略其他更重要的考量。

（四）、大數據告訴我，很少人看到最後一頁

社會科學正在成為一門真正的科學，而這門科學貨真價實的新科學將有助於改善我們的生活，這場科學革命將會以零碎的研究和發現呈現，我們慢慢會對人類心智與社會的複雜系統，得到更深入的了解。無論在什麼情況下，必須追蹤大數據的線索，並採取因應的行動。

三、啟發與心得

（一）、一舉一動都逃不出大數據的天眼

讀完這本書，讓我聯想到電影《我的少女時代》，女主角林真心說：「女生啊，是很難捉摸的~我們說沒事就是有事，沒關係~就是有關係」¹。不管男生或是女生我想大家都有遇過這種經驗，其實都是我們用說反話來掩飾自己，這也是經過大數據累積測量下的真相，就像是書中提到作者的外婆，以自己廣大的閱歷針對問題提出忠告，才會獲得大家的關注和尊重，閱歷就是大數據的資料來源，以前的人也許不知道大數據的應用，也沒聽過大數據是什麼，但是他們已經在使用了，以經驗法則來使用。

再來看看這本書的封面，一對男女睡同一張床，卻是背對背滑著手

¹ 電影-《我的少女時代》。

機，搜尋著自己想了解的事情，願意對 Google 誠實卻不願意對身旁最親密的人誠實，原來大家都很虛偽，尤其在最親近的人面前也從來都不說出真心話。現在科技日新月異，我們的個資隨時都在曝光，假設當你搜尋「寶寶衣服」，然後 Google 就會跳出其他相關字樣，像是「媽媽包」、「寶寶副食品」、「娃娃車」…等等相關資訊，Google 預測你即將要升格當媽媽，或是你家已經有小寶寶，這些都是大部分的人家有小寶寶會搜尋的相關資訊，它預測你也會需要，當然我們也會點進去看看，然後不知不覺就曝露了越來越多自身相關的訊息，除了讓 Google 預測上述幾項訊息之外，也預測這類人也許會需要「尿布大降價」、「買奶粉送玩具」、「寶寶遊戲床」…等等，凡你搜尋過的訊息都會留下數位足跡及記錄，會列入下次其他人搜尋同樣資訊的參考資料，因此 Google 會收集到更多資訊，累積成大資料庫，然後量化成數據，以便下次呈現更多元的類似訊息，提供搜尋者更多選擇。

(二)、數據翻轉謊言

挑選這本書的時候，被它的標題深深吸引著，感覺這本書是探討人性，用數據揭開人性的神秘面紗，也就是謊言背後著真相是什麼，書中提到：「人們謊報在回家途中喝了多少酒，謊稱自己多常上健身房和新鞋子買多少錢，就連沒看過的書也說自己看過。……人們對朋友說謊、對老闆說謊、對子女說謊、對父母

說謊、對醫生說謊、對老公說謊、對老婆說謊，也對自己說謊。」，
為什麼我們會這樣說？又為什麼要這樣做？作者證明大數據提供一種前所未有的新穎方式，讓我們窺探人們的內心真實世界，因為人們透過鍵盤才會卸下防備，在無意中悄悄地坦承千奇百怪的事情。

西格蒙德·佛洛伊德（Sigmund Freud）精神分析學的創始人，被稱為「維也納第一精神分析學派」，他的大量研究強調，人的意識層面負責理性的思維與情緒，而潛意識層面是非理性的大本營²。我們應該都有經驗，在鍵盤打字的時候，經常把某些字詞打錯，像是「我」打成「舞」，難道潛意識表現的是想「跳舞」？；把「姓名」打成「性名」，難道潛意識表現的是跟「性」有關的事情？；我們會說我想表示的意思當然不是這樣啊，只是不小心打錯字，為什麼要放大檢視，每個人都會有打錯字的時候，無須在意才對，但是佛洛伊德卻認為這跟我們的潛意識有非常大的相關，套一句俗語就是：「日有所思，夜有所想」，只是我們不願意去承認，覺得那是難以啟齒、羞愧的事情，與我們從小受教育正大光明的性格有違，所以選擇說謊、不承認，以掩飾內心不為人知的祕密天堂，保有那一份屬於自己的私樂園。

而當你利用 Goole 搜尋時，就揭穿你真正在意的情事，網路正努力

²維基百科-西格蒙德·佛洛伊德（Sigmund Freud）。

收集這部分有關的訊息，拼湊出真相，有了這些數據做後盾，我想真相有 90% 會是正確的，也就是有 90% 是謊言，正好呼應《注意力商人》（The Attention Merchants）：「仰賴大數據迅速拆穿我們自以為文明的假象。一本讓人既著迷又震驚，時而駭人聽聞的傑作。最棒的是，讓真相一覽無遺。」³

（三）、讀數據的驚人成果-汽車貨運調查

「汽車貨運調查主要目的係蒐集臺灣地區自用與營業貨車運送商品之種類、流向、流量及運費等動態資料，營業貨車部分：包括公司基本資料、車輛數、牌照號碼、車種、車重、載重、總重及低溫商品之運費收入占本車總運費收入比率、裝卸貨物狀況、日期、訖點之裝卸商品名稱、起點、訖點、單程區間公里、行車總次數、訖點之裝卸貨總重量、貨櫃個數、運費總收入及商品來源別等。以供研訂汽車貨運管理措施及整體運輸計畫之用，並做為編製產業關聯表之重要參據。」⁴

利用全台各個貨運業者填報上來的數據資料，累積成資料庫，每台車都是資料庫的製造商，創造出新資料，作為國家整體產業的規劃依據，數據資料已成為「隱形資產」，價值驚人。日後隨著科技的進步，每台車都安裝一台 GPS，記錄每台車的行駛軌跡，形成一個母體，直接經由

³ 注意力商人：他們如何操弄人心？揭密媒體、廣告、群眾的角力戰；原文作者：Tim Wu；作者：吳修銘。

⁴ 交通部統計處汽車貨運報告 108 年 6 月

後端電腦運用、分析，節省人力上網填報時間及成本。

四、結語：數據不是萬能，謊言卻是解藥

大數據如此好用，卻不是萬能的，無法幫你找到另一半，也無法幫你變成大富翁，一味的鑽研數字可能會使你忘記更重要的東西，讓你損失慘重，這告訴我們並不是所有的東西都可以用數據來衡量，例如：情感…等等，所以大數據不是全能的東西，它是有優點也有缺點的解答工具。

歌手林宥嘉有一首歌《說謊》，歌詞：「…我沒有說謊，我何必說謊，你懂我的，我對你從來就不會假裝，我哪有說謊，請別以為你有多難忘，笑是真的不是我逞強，…」⁵。這首歌從頭到尾都繞著說謊，他說沒說謊，實際上都是在說謊，為了自尊說謊，就是讓女主角以為自己已經忘了過去，但實際上他忘不了，最後崩潰。有時候謊言是一道美麗的彩虹，是給人希望的太陽，是自我安慰的解藥，不道破假像是一種修養，真相就留給大數據及時間去證明吧！

五、參考文獻

- (一)、電影-我的少女時代。
- (二)、維基百科-西格蒙德·佛洛伊德 (Sigmund Freud)。
- (三)、注意力商人：他們如何操弄人心？揭密媒體、廣告、群眾的角

⁵林宥嘉-第二章個人專輯《感官/世界》。

力戰。

(四)、林宥嘉-第二章個人專輯《感官/世界》。

(五)、《商業週刊》第 1410 期-《經濟學人》它，比你更了解你自己！一次看懂大數據的威力。