

資料分析通識

數據交點 | 郭耀仁 yaojenkuo@datainpoint.com

三個資料分析的重要思維

1. **雙城奇謀**，商業與資料導向分析的模糊地帶。
2. **泛用迷思**，沒有一體適用所有應用場景的分析框架。
3. **山羊過橋**，總有一方要弭平商業與資料導向間的分歧。

雙城奇謀



來源: <http://gph.is/1WbH9RY>

什麼是分析

將某個有興趣的問題拆分為數個小問題，研究小問題與有興趣的問題之間的關係。

平鋪直敘的說法

撇除所有華麗的詞彙，「資料分析」是透過以邏輯、事實以及量化為基準來解決問題的方法。

為什麼要透過資料分析解決問題

- 分析是連接問題與行動計畫之間的橋樑。
- 商業問題 -> ~~資料分析~~ -> 行動計畫 -> 評估 -> ...重複循環

常見的十個（商業）問題

- 進入一個新的市場
- 產業分析
- 併購
- 新產品策略
- 定價策略

來源：<https://www.amazon.com/Case-Point-11-Interview-Preparation/dp/0986370762>

常見的十個（商業）問題（續）

- 成長策略
- 新事業
- 競爭策略
- 增加利潤
 - 提升營業收入
 - 減少營業成本
- 重組

來源：<https://www.amazon.com/Case-Point-11-Interview-Preparation/dp/0986370762>

因應而生的分析框架

- 增進分析的效率、品質、可靠性與強韌性。
- 實踐 DRY(Don't Repeat Yourself)準則。

太多的分析框架也是問題所在



來源: <http://gph.is/18hNtOr>

雙城奇謀：商業導向分析師

商業導向分析師偏好的分析框架：

- 常春藤系統（3Cs、4Ps、波特五力）。
- 企管顧問公司系統（BCG 矩陣、麥肯錫七步驟）。
- 損益表分析。
- ...等。

雙城奇謀：資料導向分析師

資料導向分析師偏好的分析框架：

- Gartner 分析遞增模型。
- 資料科學金字塔需求。
- 資料科學專案流程。

常春藤系統：3Cs

- Customer
- Competitor
- Company

常春藤系統： 4Ps

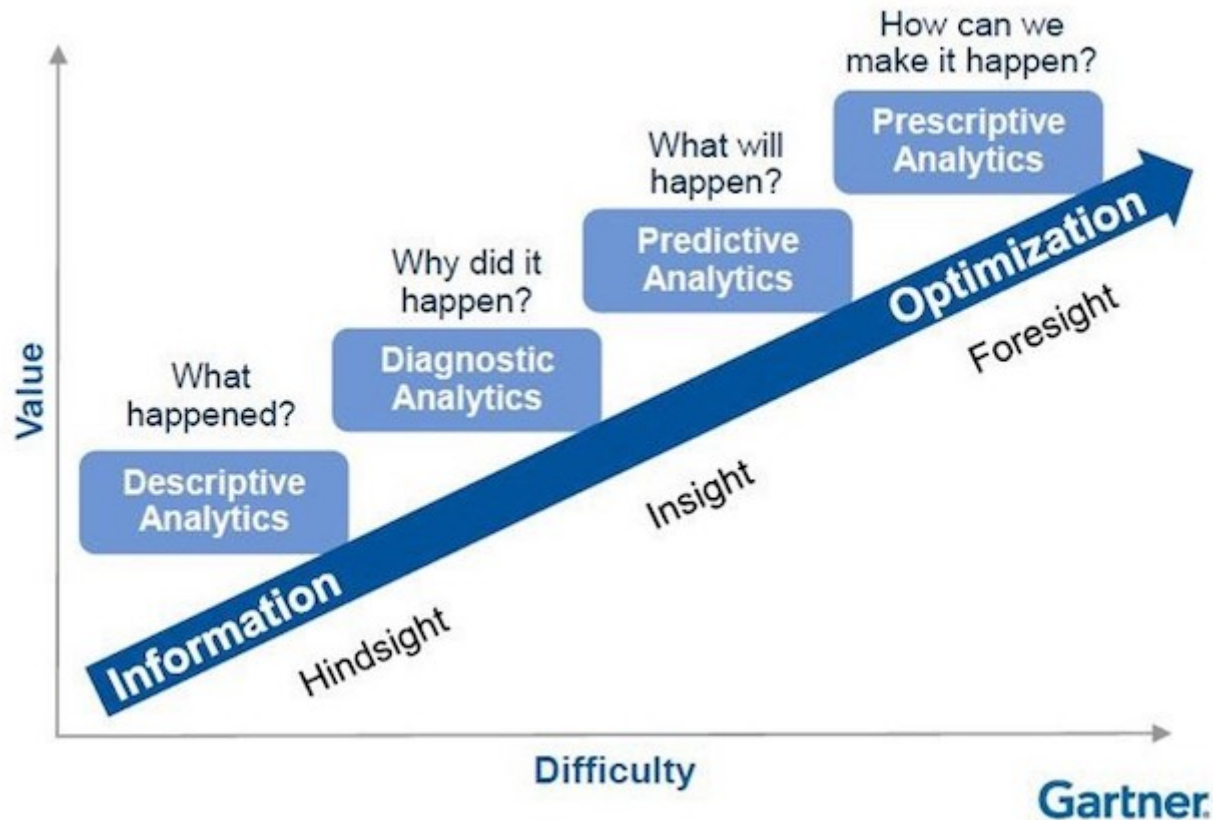
- Product
- Price
- Place
- Promotion

企管顧問公司系統：麥肯錫七步驟

1. 定義問題。
2. 拆解問題。
3. 優先排序。
4. 議題分析。
5. 整合。
6. 故事敘述。
7. 簡報。

...重複循環。

Gartner 分析遞增模型



來源: <https://www.gartner.com/en/documents/1964015/best-practices-in-analytics-integrating-analytical-capab>

資料科學金字塔需求

THE DATA SCIENCE HIERARCHY OF NEEDS

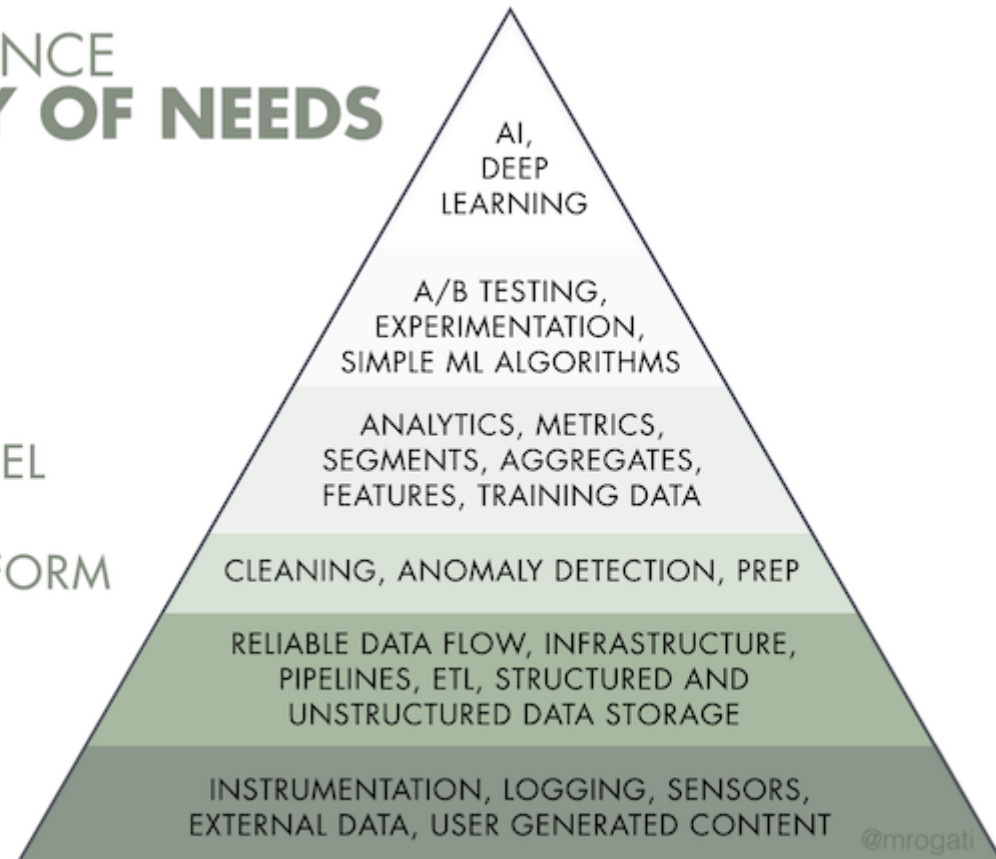
LEARN/OPTIMIZE

AGGREGATE/LABEL

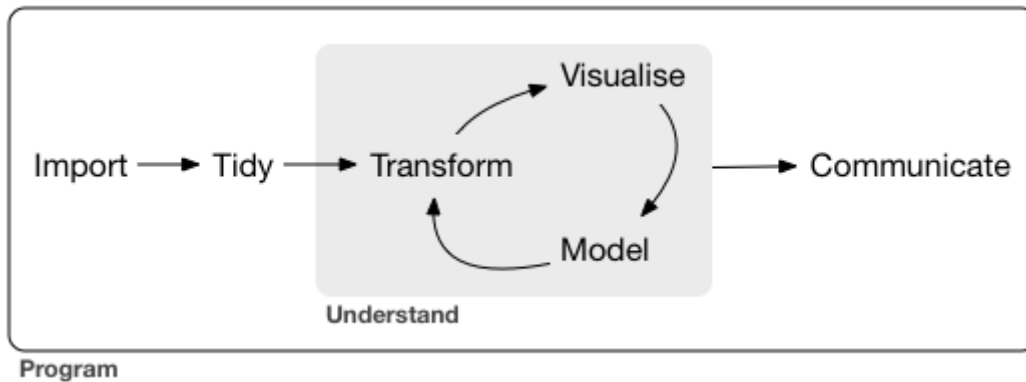
EXPLORE/TRANSFORM

MOVE/STORE

COLLECT



資料科學專案流程



來源: <https://r4ds.had.co.nz/introduction.html>

不同的專注

- 商業導向專注在推論邏輯。
- 資料導向專注在資訊價值。

小結：我們從「雙城奇謀」得到什麼？

- 當我們談到「分析」的時候，隨著聽眾的背景不同將會產生一定程度的模糊性。
- 我們談到了「兩種口味」的分析：商業導向以及資料導向。
- 基於兩種導向各自衍生出多樣化的分析框架。

泛用迷思

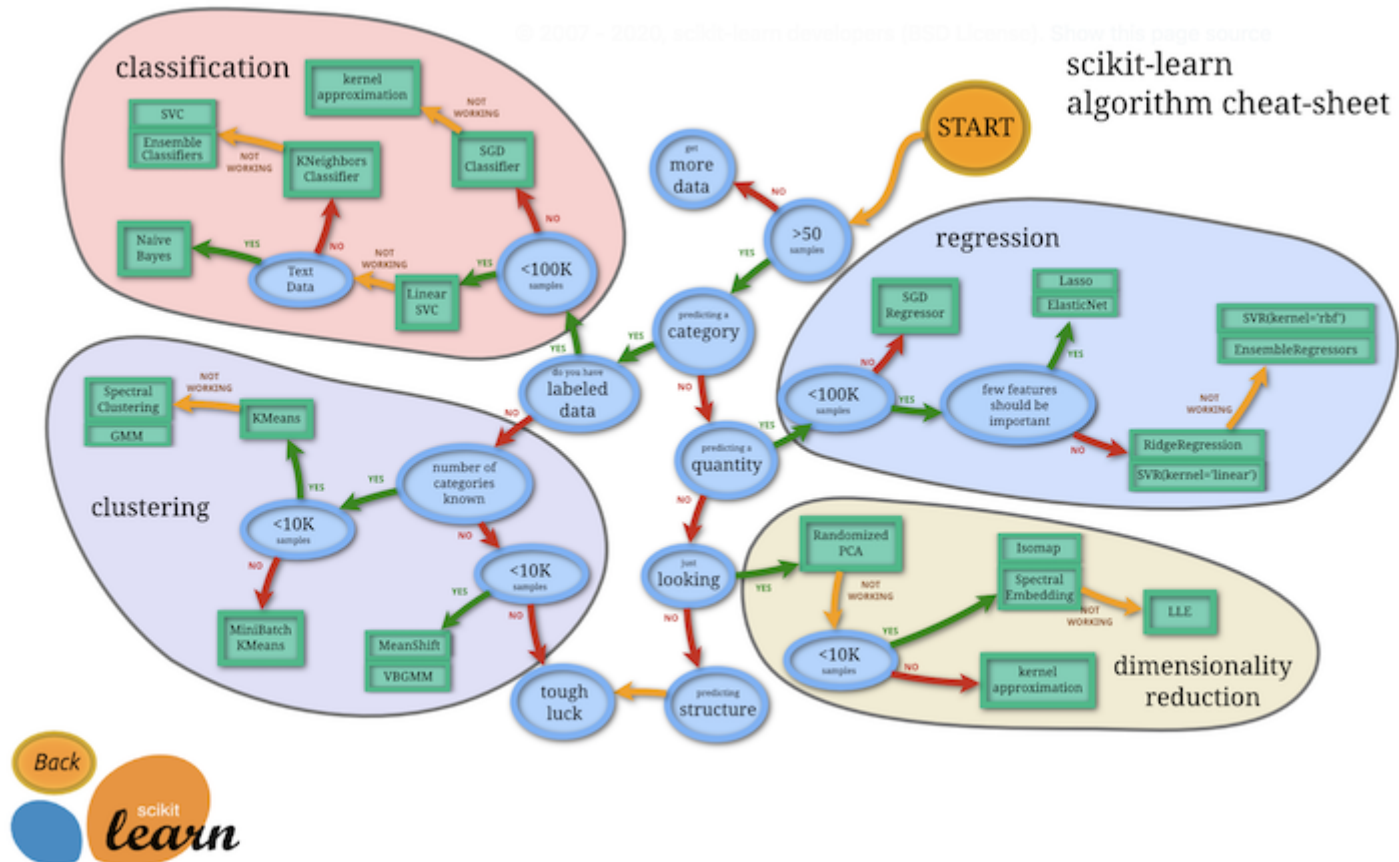


來源：<http://gph.is/1maf0K9>

童話故事

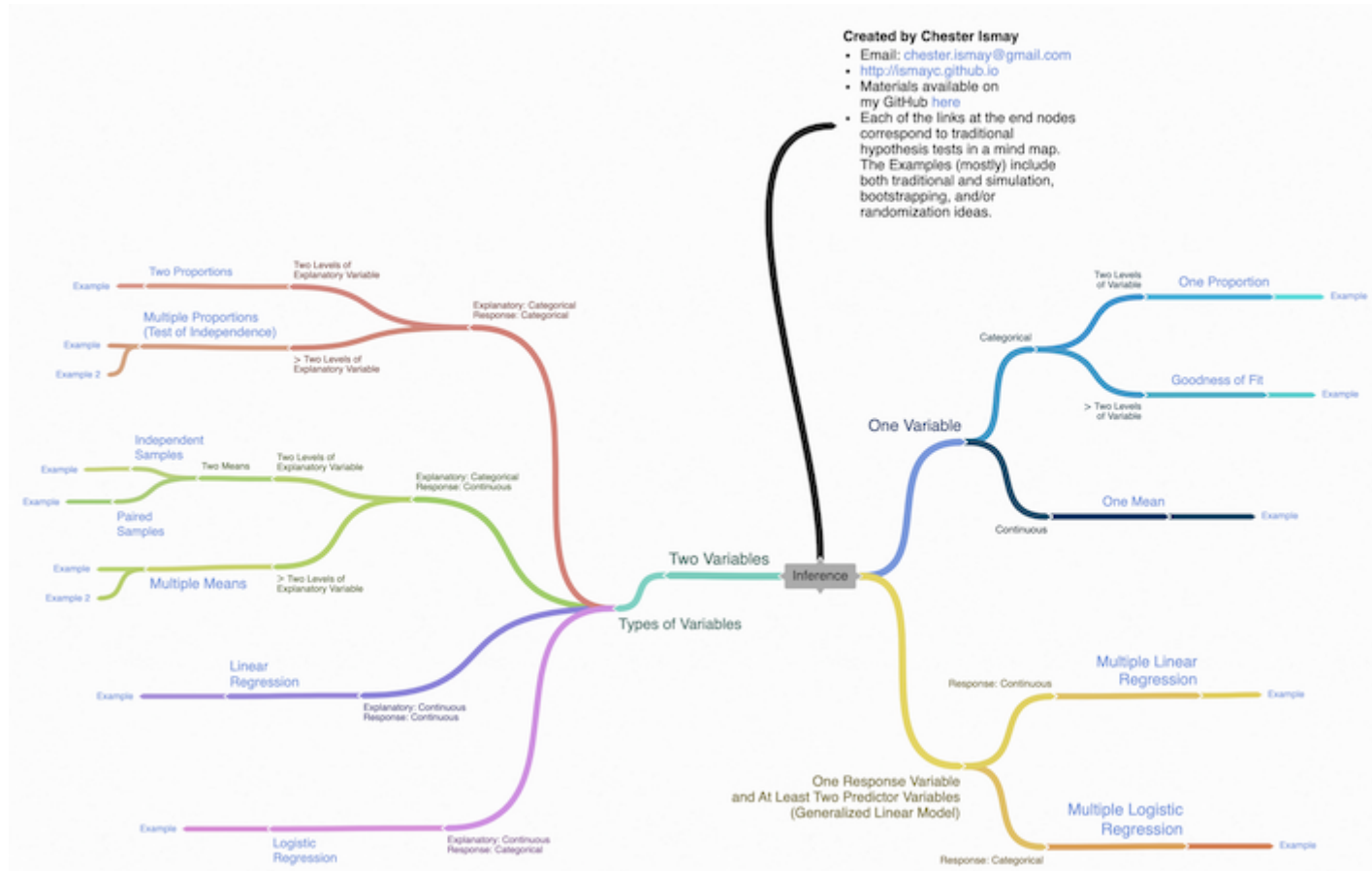
有一個泛用的分析框架可以套用在所有的專案中。

SCIKIT-LEARN: 選擇不同的估算器



來源: https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

CHESTER ISMAY: 統計推論心智圖



來源: <https://coggle.it/diagram/Vxlydu1akQFeqo6-/t/inference>

有眾多的估算器與統計推論模型

我們沒有自信說「最佳模型」，但是我們可以說明在某些應用場景下，特定模型比其他的模型表現更好。

有眾多的框架

我們沒有自信說「最佳框架」，但是我們可以說明在某些應用場景下，特定框架比其他的框架更能幫助分析師解決問題。

泛用的框架

- 資料導向框架：麥肯錫七步驟
- 資料導向框架：Gartner 分析遞增模型。

簡化的麥肯錫七步驟

1. 定義問題。
2. 拆解問題。
3. 優先排序。
4. ~~議題分析。~~
5. 整合。
6. ~~故事敘述。~~
7. ~~簡報。~~

...重複循環。

好的「定義問題」

- 明確。
- 可爭論。
- 專注在決策者的行動。

輔助定義問題的工作表

PROBLEM STATEMENT WORKSHEET	
Basic question to be resolved Define what must be resolved to deliver impact for the client. The question should be SMART. The more specific the statement the better.	
1 Perspective/context Spells out the situation and complication facing the client. (e.g. trends in the industry, relative position in the industry etc)	2 Decision makers/stakeholders Identify primary decision makers (e.g. CEO, Division Manager) as well as internal and external parties that can support/derail the project (e.g. shareholders, regulatory, authorities, SBU managers)
3 Criteria for success Defines success for the project. Must be shared by client and engagement team. Should include relevant quantitative and qualitative measures. (e.g. visible improvements, effect on staff mindset, market growth)	4 Scope of solution space Indicate what will and will not covered by the engagement team, including accuracy and quality expectation
5 Constraints within solution space Defines the limits of the set of solutions that can be considered (e.g. team can only consider organic growth and not inorganic)	6 Key sources of insights Research on where expertise, knowledge, engagement methodology exist both internally (within the firm) and externally (client, specialist firms etc) (e.g. practice databases)

好的「拆解問題」

- 善用邏輯樹。
 - 議題樹。
 - 假說推論樹。
- 互斥周延法則（MECE）。

議題樹：前向推理

- 如何存更多錢？
 - 增加收入。
 - 更多的營業收入。
 - 升遷。
 - 跳槽。
 - 業績獎金。
 - 加班費。
 - 更多的投資收入。
 - 調整投資組合。
 - 縮減支出。
 - 買即期品。
 - 買快時尚。
 - 現金回饋計畫。

假說推論樹：後向推理

- 透過調整投資組合存更多錢。
 - 沒有業績獎金與加班費
 - 短期沒有升遷或跳槽規劃。
 - 不想改變消費習慣。
 - 投資過於保守，沒有槓桿。
 - 沒有空頭部位，無法在熊市賺錢。

「拆解問題」的秘訣

There is only one way to eat an elephant: a bite at a time.

Desmond Tutu

好的「優先排序」

同時將財務效果與方案可行性納入考量。

建立一個「效果-可行性」矩陣

Analytics use cases should be prioritized based on feasibility and impact.

Step 1: Create a list of use cases.

Sample list for consumer-packaged-goods company

Sales/customer relationship management (CRM)

1. Overall brand management
2. Overall campaign management
3. 360° view of shopper
4. Targeted acquisition campaigns
5. Real-time image advertising (awareness)
6. Retargeting campaign

Marketing

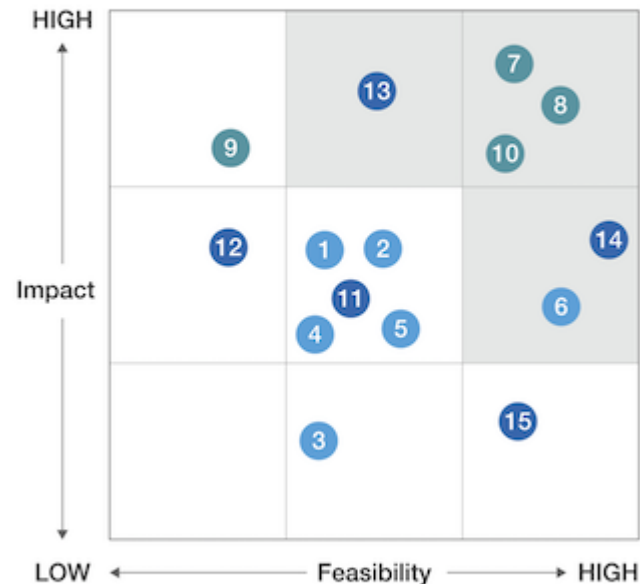
7. Optimization of spend across media
8. Optimization of spend within digital media
9. Digital attribution modeling
10. Performance advertising (sales)

Innovation

11. Consumer insights (social listening/sentiment analysis)
12. New product success (predictive behavior model)
13. Product customization at scale
14. Open innovation on promotion mechanisms
15. New digital sales models

Step 2: Prioritize them.

Sample impact vs feasibility matrix



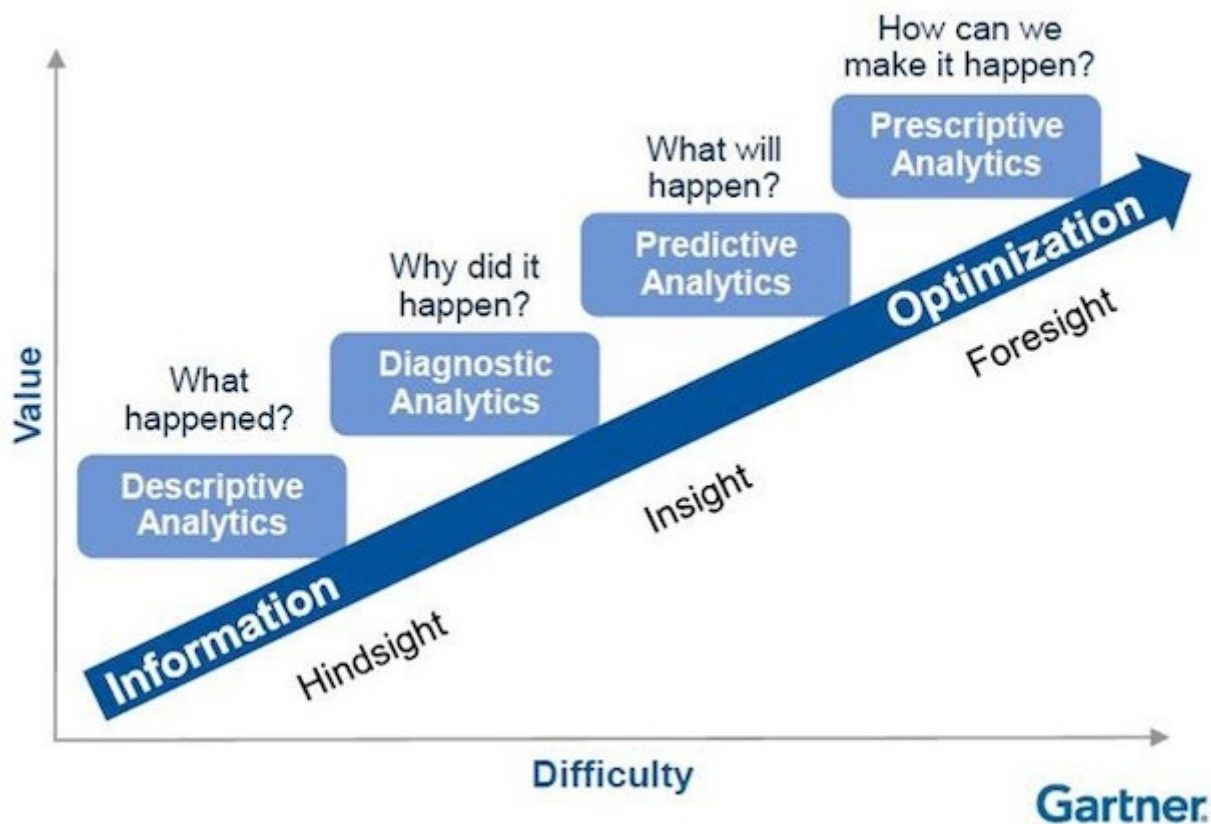
McKinsey&Company

來源: [Ten red flags signaling your analytics program will fail](#)

好的「整合」

- 連結假說與邏輯。
- 標註必要資訊。
- 標註責任與時程。

Gartner 分析遞增模型



來源: <https://www.gartner.com/en/documents/1964015/best-practices-in-analytics-integrating-analytical-capab>

定義：敘述型分析

用來回答「發生了什麼事件？」的分析型態，利用商業智能軟體、視覺化、圖表與描述等手法回答。

定義：診斷型分析

用來回答「為何會發生這樣的事件？」的分析型態，利用組成下鑽、探索性分析、資料探勘與相關性分析等手法回答。

定義：預測型分析

用來回答「如何預測何時會發生這樣的事件？」的分析型態，利用時間序列分析、機器學習與神經網路等手法回答。

定義：處方型分析

用來回答「如何讓事件發生？」的分析型態，利用時間序列分析、機器學習與神經網路搭配模擬與實驗等手法回答。

Gartner 分析遞增模型的優點

- 與傳統的 **5W2H** 記者報導框架相似，容易理解。
- 直觀的概念，先學會爬、再學會走、最後是跑。

Gartner 分析遞增模型的缺點

- 「遞增」隱喻了導入的優先順序。
- 兩軸「價值」與「難度」容易有誤導的嫌疑。
- 更高層級的分析並不能保證帶來更高價值。

小結：我們從「泛用迷思」得到什麼？

- 沒有一體適用的分析框架，但在某些應用場景中也許有較為泛用的選擇。
- 更瞭解麥肯錫七步驟與 Gartner 分析遞增模型。

山羊過橋



來源：<http://gph.is/1NTV4c8>

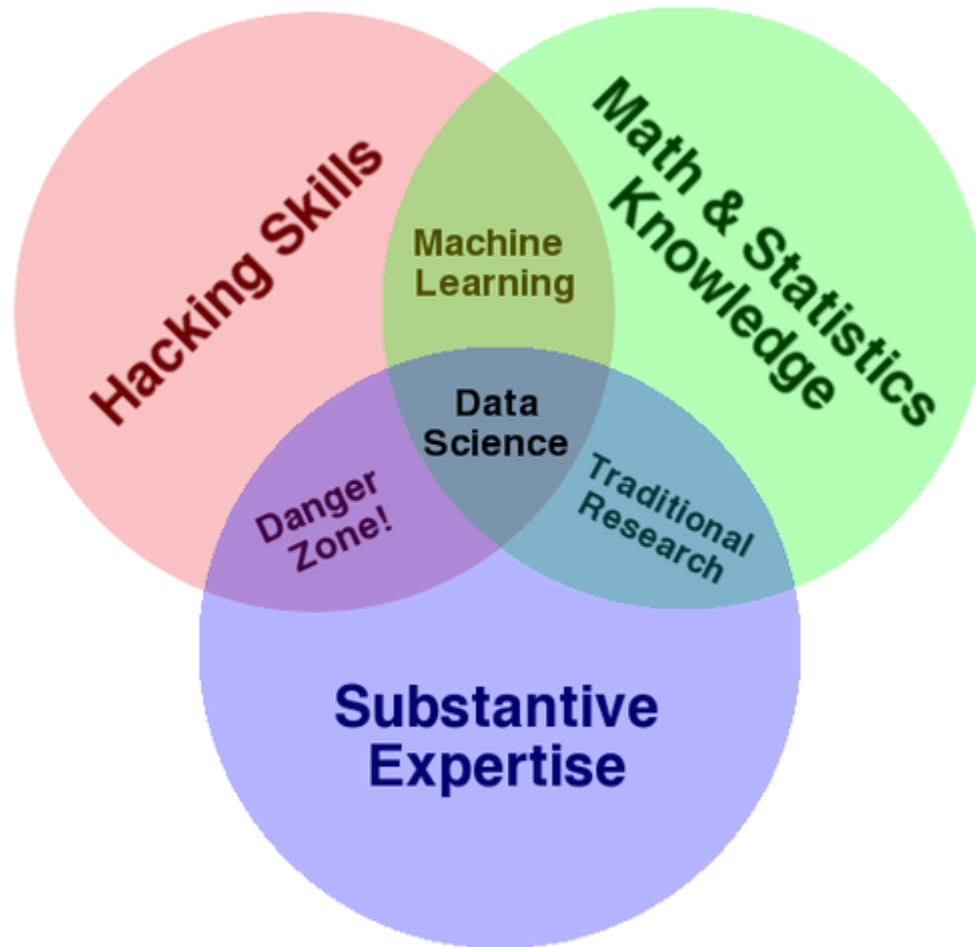
行動方案

我們現在已經清楚瞭解關於商業與資料導向分析師的分歧，是時候將分歧弭平了。

哪個比較困難？

- 商業導向的分析師理解資料科學。
- 資料導向的分析師理解商業問題。

常見的資料科學文氏圖



來源: <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>

從較為泛用的框架中去蕪存菁

- 簡化的麥肯錫七步驟。
- Gartner 分析遞增模型中的優點。

混成分析框架

- 「商業導向」的分析框架。
- 「資料導向」的分析框架。

BE WATER MY FRIEND

Be formless, shapeless, like water.

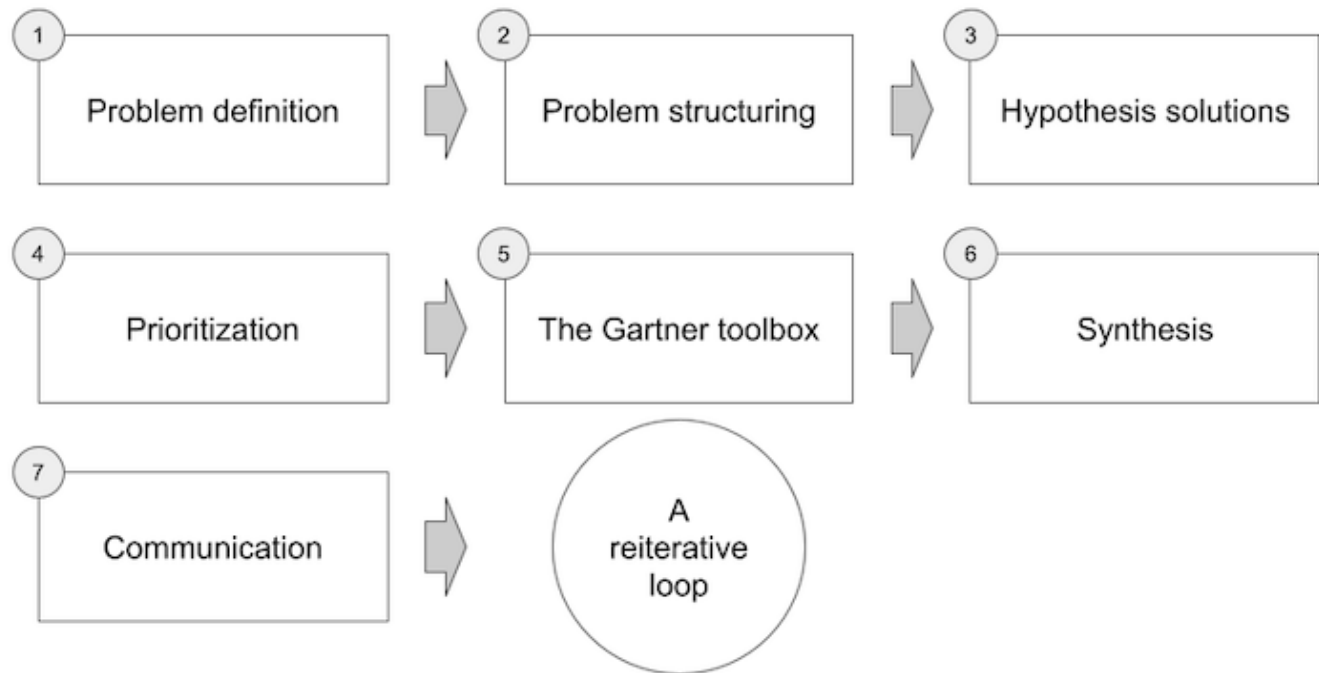
來源: <https://youtu.be/uk1lzkH-e4U?t=945>

混成分析框架：如何橋接商業問題與行動方案

1. 定義問題。
2. 拆解問題。
3. 建立假說。
4. 優先排序。
5. Gartner 分析遞增模型。
6. 整合。
7. 溝通。

...重複循環。

畫得不那麼漂亮的混成分析框架示意



好的「資料分析師」

- 對於技術有飢渴感（tech-savvy）。
- 對於商業有熱情。

由資料導向的分析師跨出弭平分歧的那一步

- 在「定義問題」環節投注更多的努力。
- 學習「優先排序」以及「整合」的軟實力（從專案經理身上學！）。
- 在「溝通」環節投注更多的努力。

如何在「定義問題」環節投注更多的努力

- 辨識利益關係人：決策者、關係人以及自身。
- 把利益關係人當作合作夥伴，先不要有太大的防衛心。
- 經營出熱情、客觀與溝通無礙的關係。

YOU HAVE GOT TO LISTEN

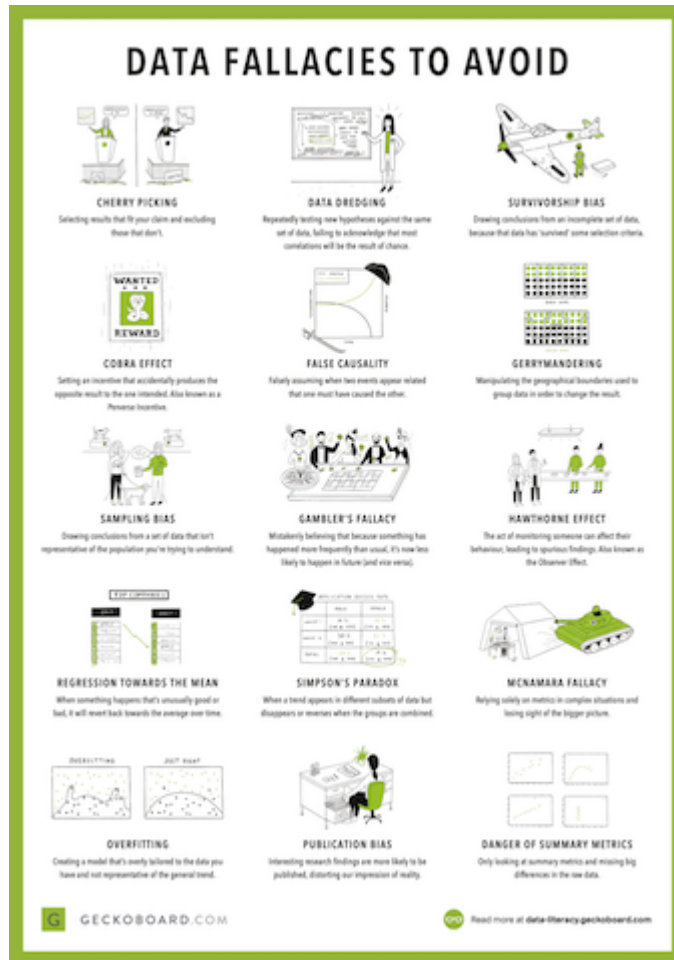


來源: <https://gph.is/g/EGyoXQ3>

學習「優先排序」以及「整合」的軟實力

- 避免犯下常見的資料謬論。
- 從「混淆矩陣」向「成本效益矩陣」邁進。
- 80/20 法則。
- 有效的行動方案要納入責任歸屬、時程與明確步驟。

避免犯下常見的資料謬論



來源: <https://www.geckoboard.com/best-practice/statistical-fallacies/>

成本效益矩陣

		Actual	
		p	n
Predicted	Y	$b(Y,p)$	$c(Y,n)$
	N	$c(N,p)$	$b(N,n)$

來源: <https://www.amazon.com/Data-Science-Business-Data-Analytic-Thinking/dp/1449361323>

在「溝通」環節投注更多的努力

- 試著將「邏輯」保留到最後一招。
- 不要只給溝通對象（客戶）一個選項。

說服的藝術：亞里士多德框架

- Pathos
- Ethos
- Logos

小結：我們從「山羊過橋」得到什麼？

- 由好的資料分析師邁出弭平分歧的那一步。
- 好的分析師應該加強哪些能力來弭平分歧。
- 應用混成框架。