

組織內推動數據驅動決策- Data Science & Data Driven Approach

@ R Ladies Taipei 2018/04/30

Angie Hsu 徐紹婷

About Me- Angie Hsu 徐紹婷



清華大學經濟系

中山大學企業管理研究所

遠傳電信企業策略發展部數據分析師
(2015.02-現在)

R Ladies Taipei Co-Organizer 之一

R Ladies Taipei R Basic讀書會小組成員

<https://www.linkedin.com/in/angie-hsu-shaoting/>

推動數據驅動決策的3W1H

- Where?
 - 在組織內or組織外
 - 在組織裡的哪裡做
- What?
 - 解決什麼問題
 - 衡量成功的指標是什麼
- Who?
 - 團隊裡面有哪些角色
 - 真正可以決定決策是否推行的key man是誰
- How?
 - 如何開啟數據驅動決策流程
 - 如何真正發揮”Data Science”的影響力

推動數據驅動決策的3W1H: Where

- Where?
 - 在組織內還組織外
 - 在組織裡的哪裡做
- What?
 - 解決什麼問題
 - 衡量成功的指標是什麼
- Who?
 - 團隊裡面有哪些角色
 - 真正可以決定決策是否推行的key man是誰
- How?
 - 如何開啟數據驅動決策流程
 - 如何真正發揮”Data Science”的影響力

Data Science的服務提供者在組織內還組織外

- 組織外
 - 外包一部分Data相關工作給廠商
 - e.g. 購買儲存資料設備, 預測模型開發, 報表系統建置
- 組織內
 - 其他不適合外包的工作
 - 招募data engineer, data analyst, data scientist等
 - 訓練內部有領域專業的既有員工

在組織裡的哪裡做

-Data Science/ Analytics team在組織內的配置

Centralized



Decentralized



Hybrid



推動數據驅動決策的3W1H: What

- Where?
 - 在組織內還組織外
 - 在組織裡的哪裡做
- What?
 - 解決什麼問題
 - 衡量成功的指標是什麼
- Who?
 - 團隊裡面有哪些角色
 - 真正可以決定決策是否推行的key man是誰
- How?
 - 如何開啟數據驅動決策流程
 - 如何真正發揮”Data Science”的影響力

想用Data解決什麼商業問題

- 我們是為誰解決了什麼問題？
- 它會如何有效影響業務？
- 這些資料來自哪裡？輸出是什麼？它是如何被使用的？
（線上演演算法、靜態報表等）
- 這是一個cost-saving還是revenue-increasing的問題？

不同產業常見的商業問題



Retail



Manufacturing



Banking



Healthcare

不同產業常見的商業問題

- 購物籃分析
 - 通過分析使用者的購物籃，提升推薦產品的購買率
 - 影響：增加收入
- 好友推薦
 - 根據使用者的情況，給他推薦最適合的好友
 - 影響：優化產品體驗
- 風險預估
 - 給定借款人和貸款的特點，預測債務是否能得到回收？
 - 影響：管理風險
- 保修預測
 - 預測產品是否需要保修
 - 影響：優化管理
- 藥物（劑量）有效性
 - 預測不同類型、劑量的藥物對治療疾病的效果
 - 影響：提升就醫效果

Source:
<https://mp.weixin.qq.com/s/GwppobhEif3iYBbXmwAX7w>

Airbnb-- How well does NPS predict rebooking?

- NPS, or the “Net Promoter Score”, is a widely used customer loyalty.
- Ask guests “How likely are you to recommend Airbnb to a friend?”—a question called “likelihood to recommend” or LTR

How likely are you to recommend Airbnb to a friend?

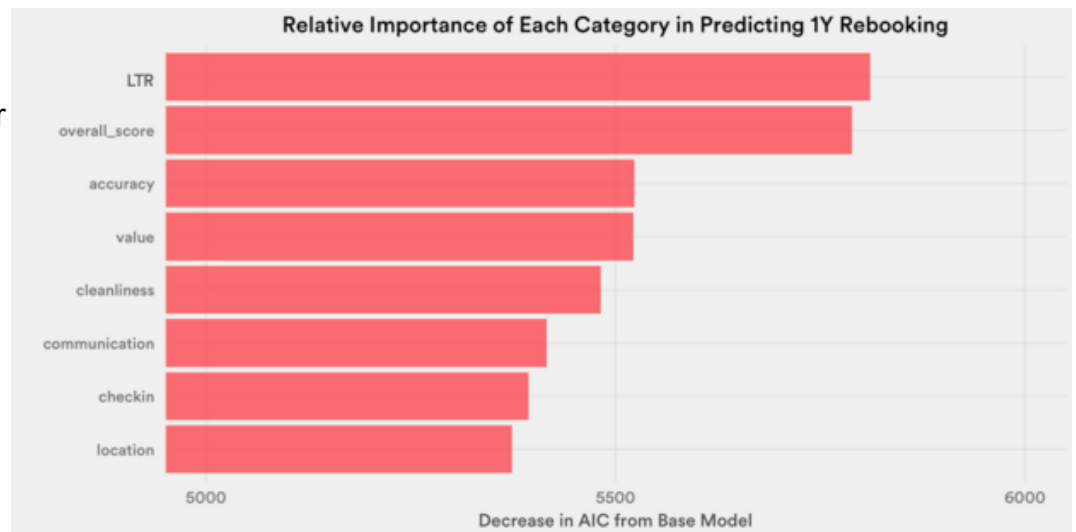
This answer will not be shared with your host or posted on your profile.

Not at all likely Extremely Likely

0 1 2 3 4 5 6 7 8 9 10

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☒

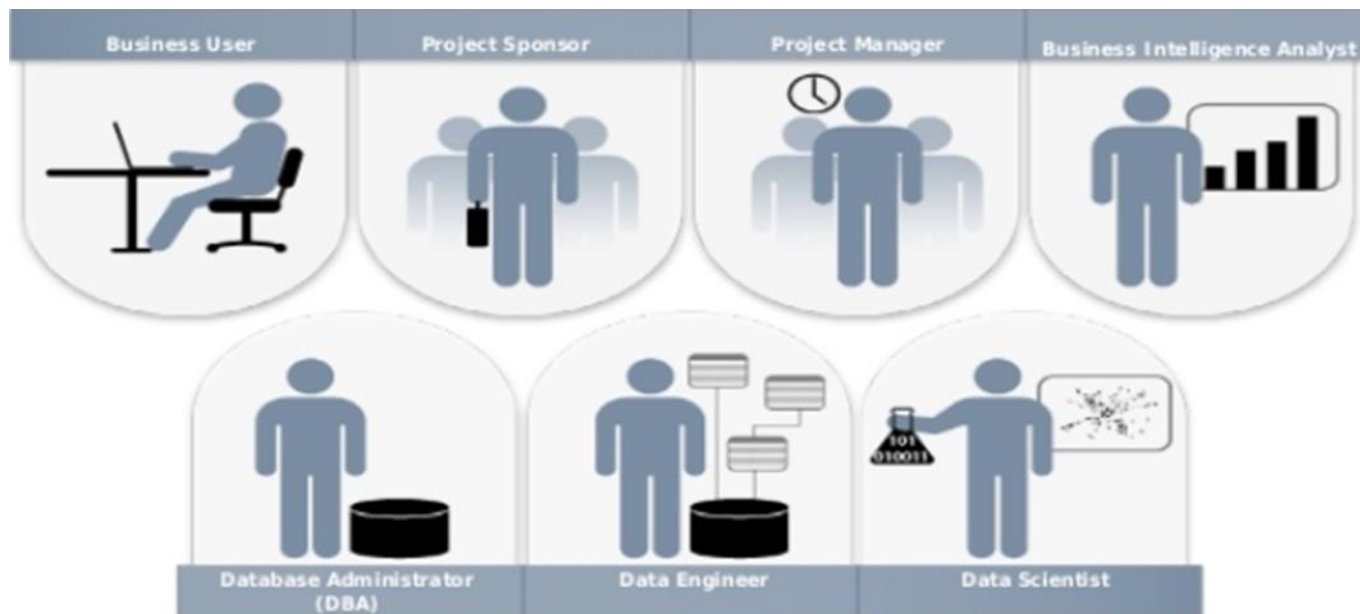
Next



推動數據驅動決策的3W1H: Who

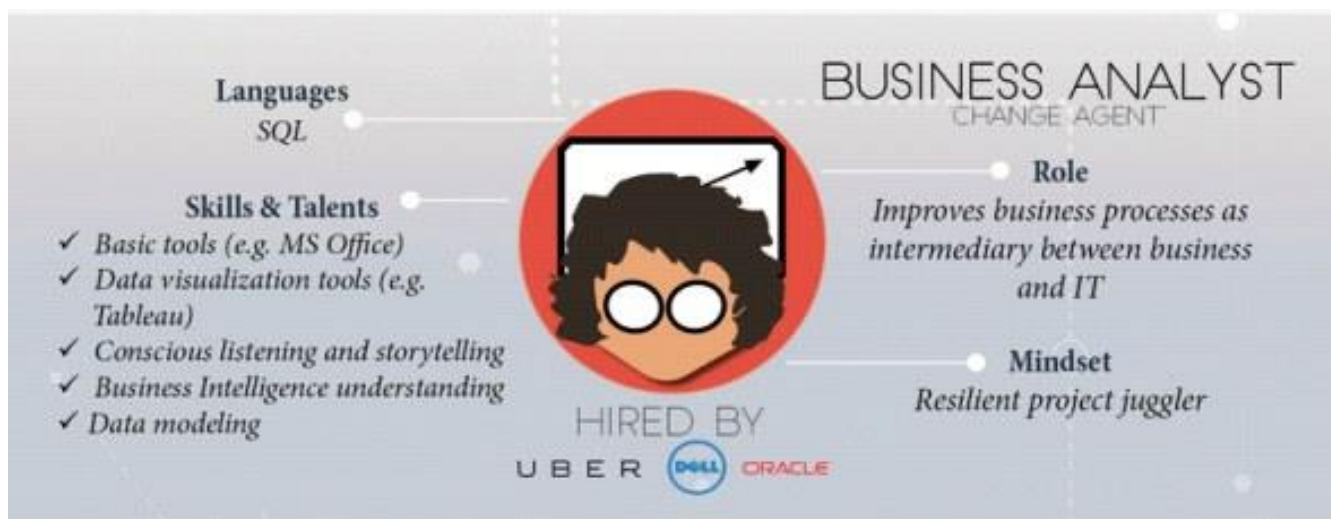
- Where?
 - 在組織內還組織外
 - 在組織裡的哪裡做
- What?
 - 解決什麼問題
 - 衡量成功的指標是什麼
- Who?
 - 裡面有哪些角色
 - 真正可以決定決策是否推行的key man是誰
- How?
 - 如何開啟數據驅動決策流程
 - 如何真正發揮”Data Science”的影響力

Data Science專案中從頭到尾會有哪些角色參與



Source: <https://www.slideshare.net/emcacademics/building-data-science-teams-31057129>

不同職能所需要的技能



推動數據驅動決策的3W1H: How

- Where?
 - 在組織內還組織外
 - 在組織裡的哪裡做
- What?
 - 解決什麼問題
 - 衡量成功的指標是什麼
- Who?
 - 團隊裡面有哪些角色
 - 真正可以決定決策是否推行的key man是誰
- How?
 - 如何開啟數據驅動決策流程
 - 如何真正發揮”Data Science”的影響力

THE DATA SCIENCE HIERARCHY OF NEEDS

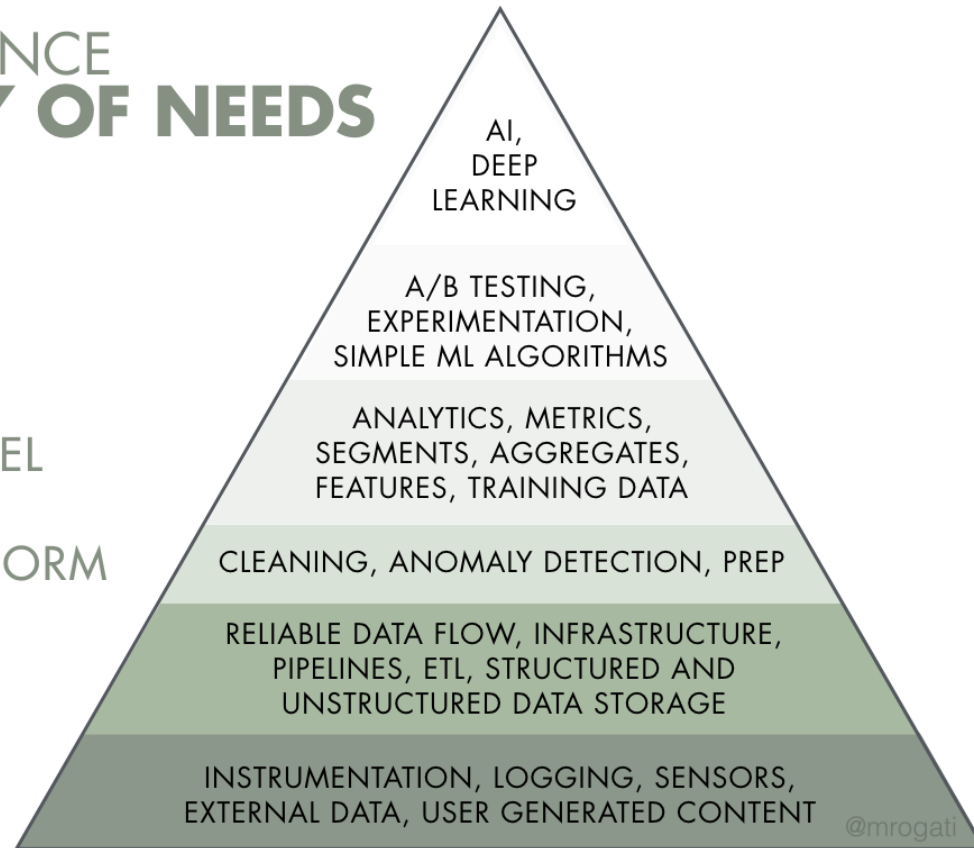
LEARN/OPTIMIZE

AGGREGATE/LABEL

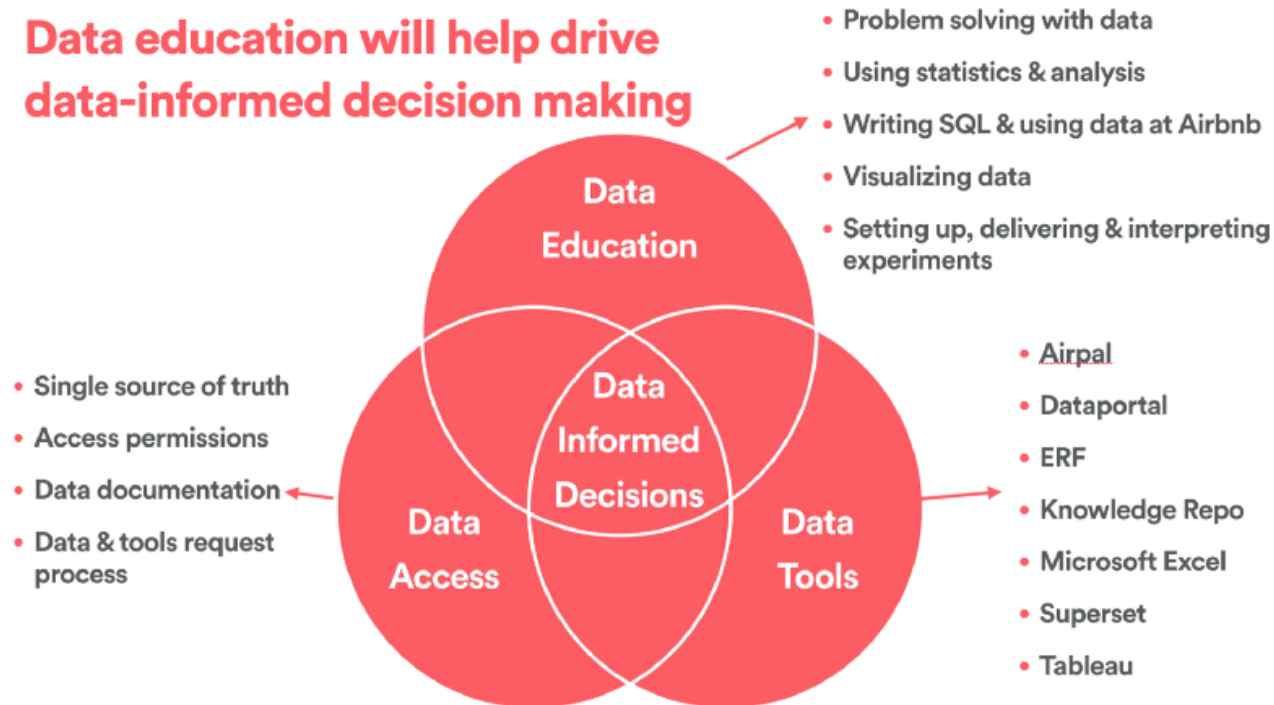
EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

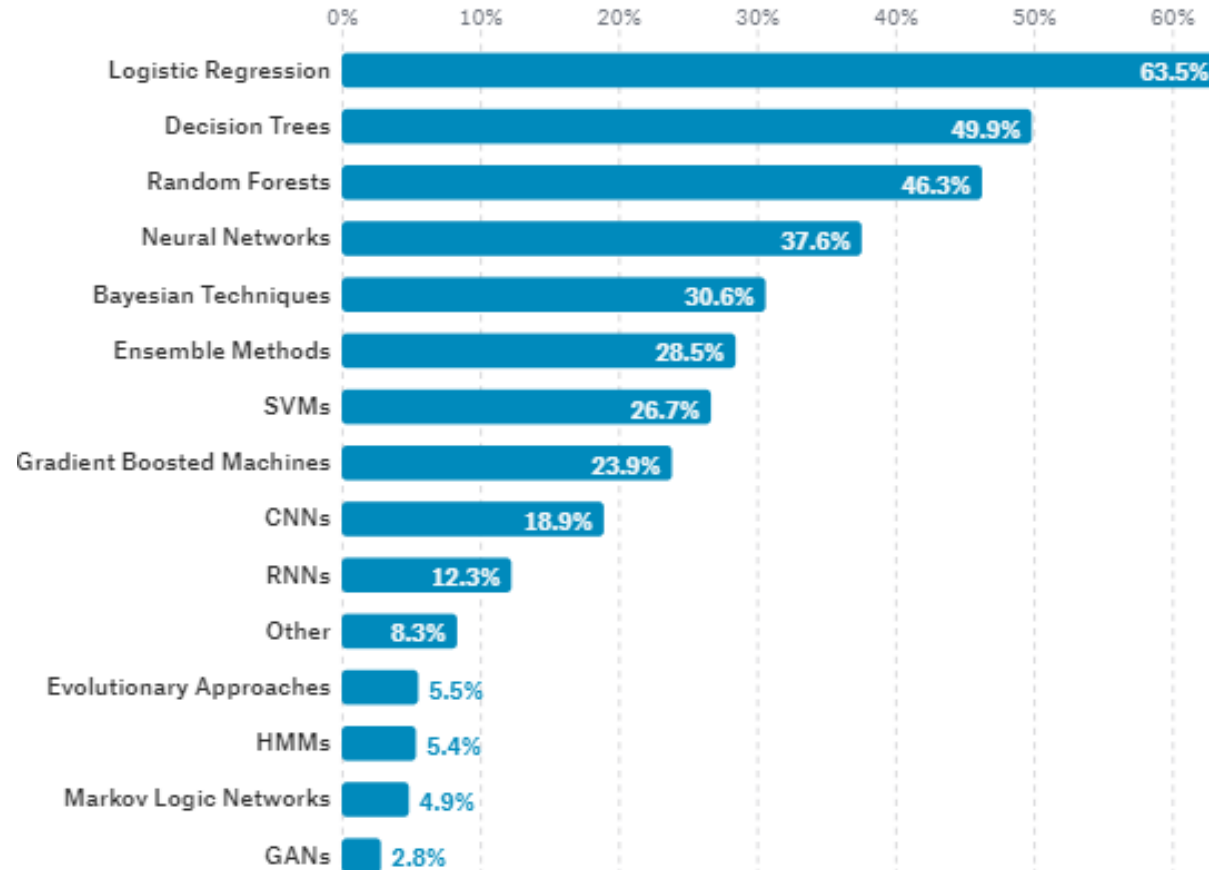


Data education will help drive data-informed decision making

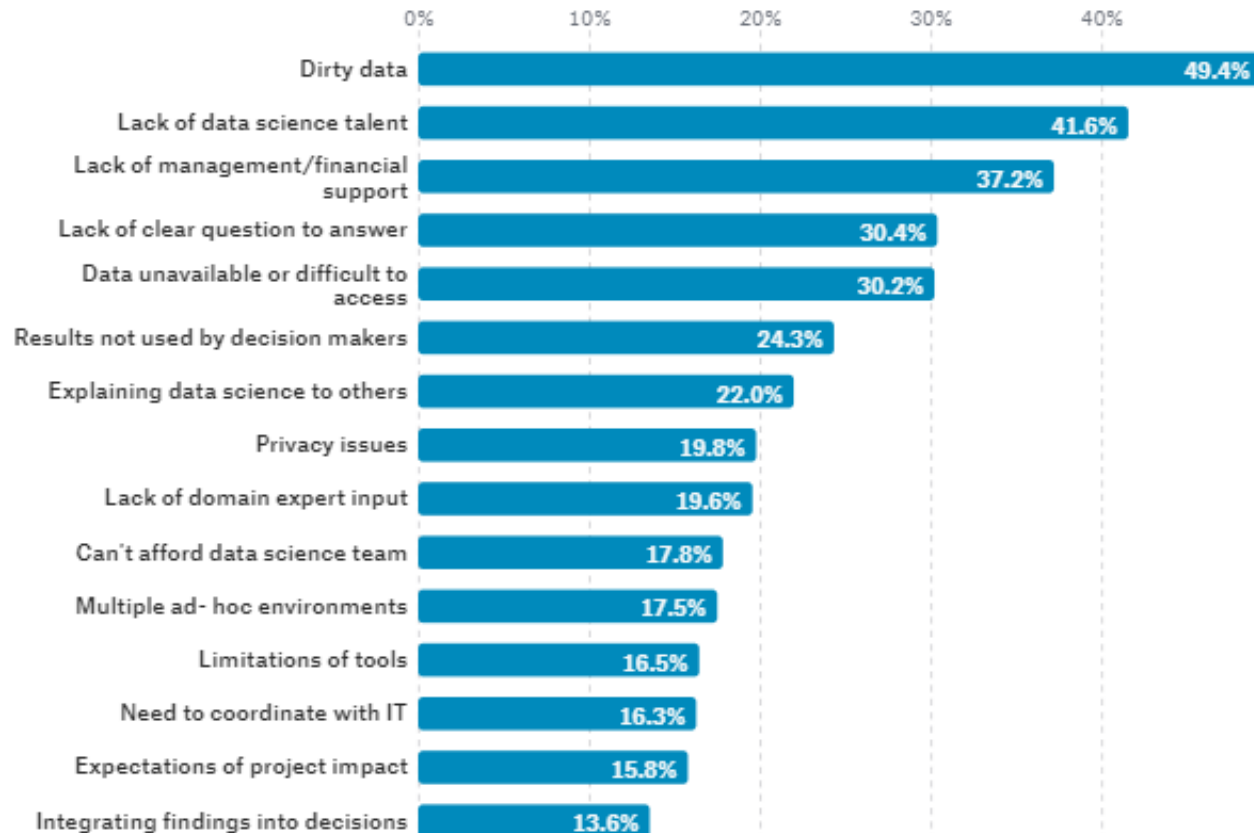


Source: <https://medium.com/airbnb-engineering/how-airbnb-democratizes-data-science-with-data-university-3eccc71e073a>

What data science methods are used at work?

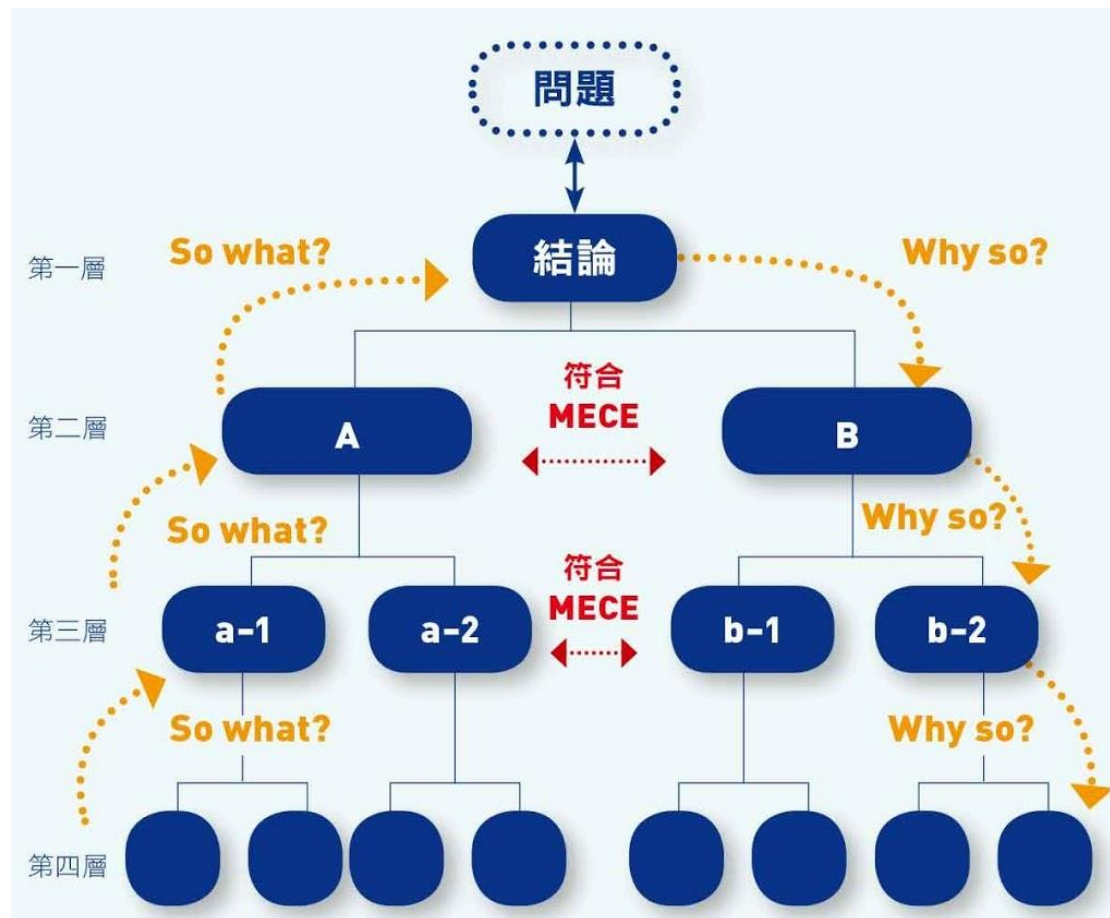


What barriers are faced at work?



如何增進溝通效率

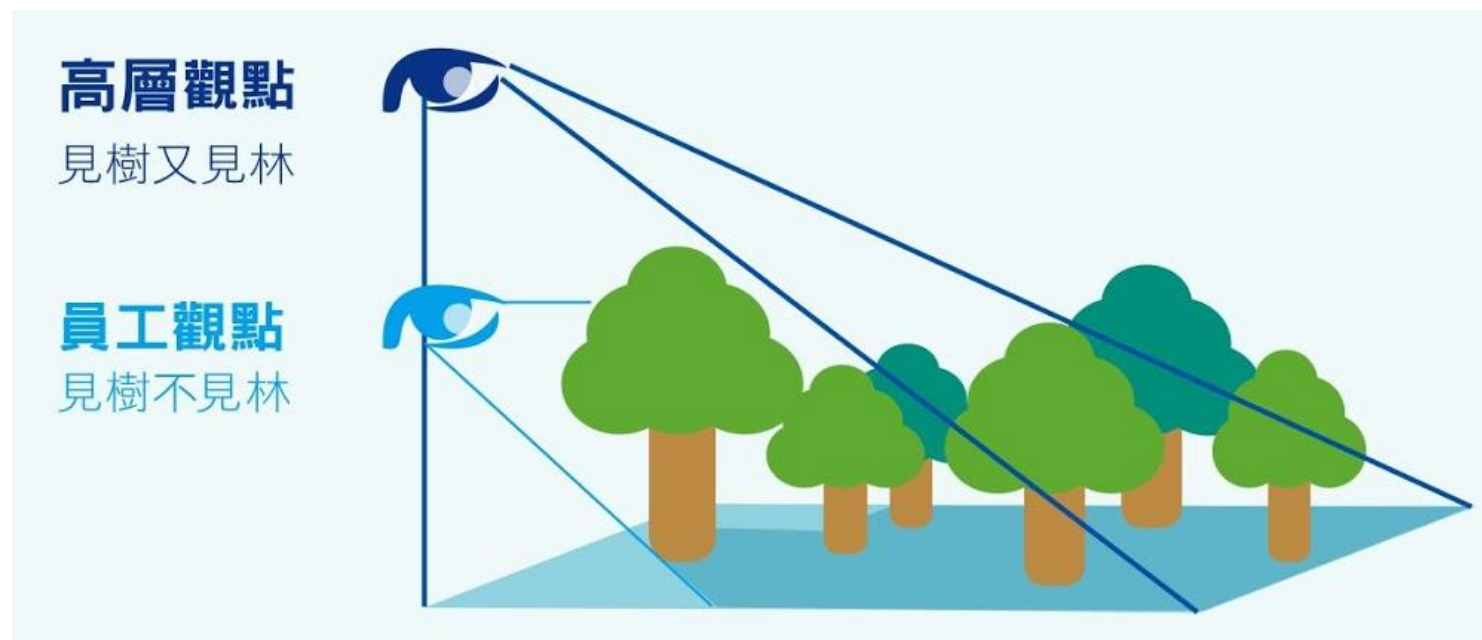
參考麥肯錫-思考架構



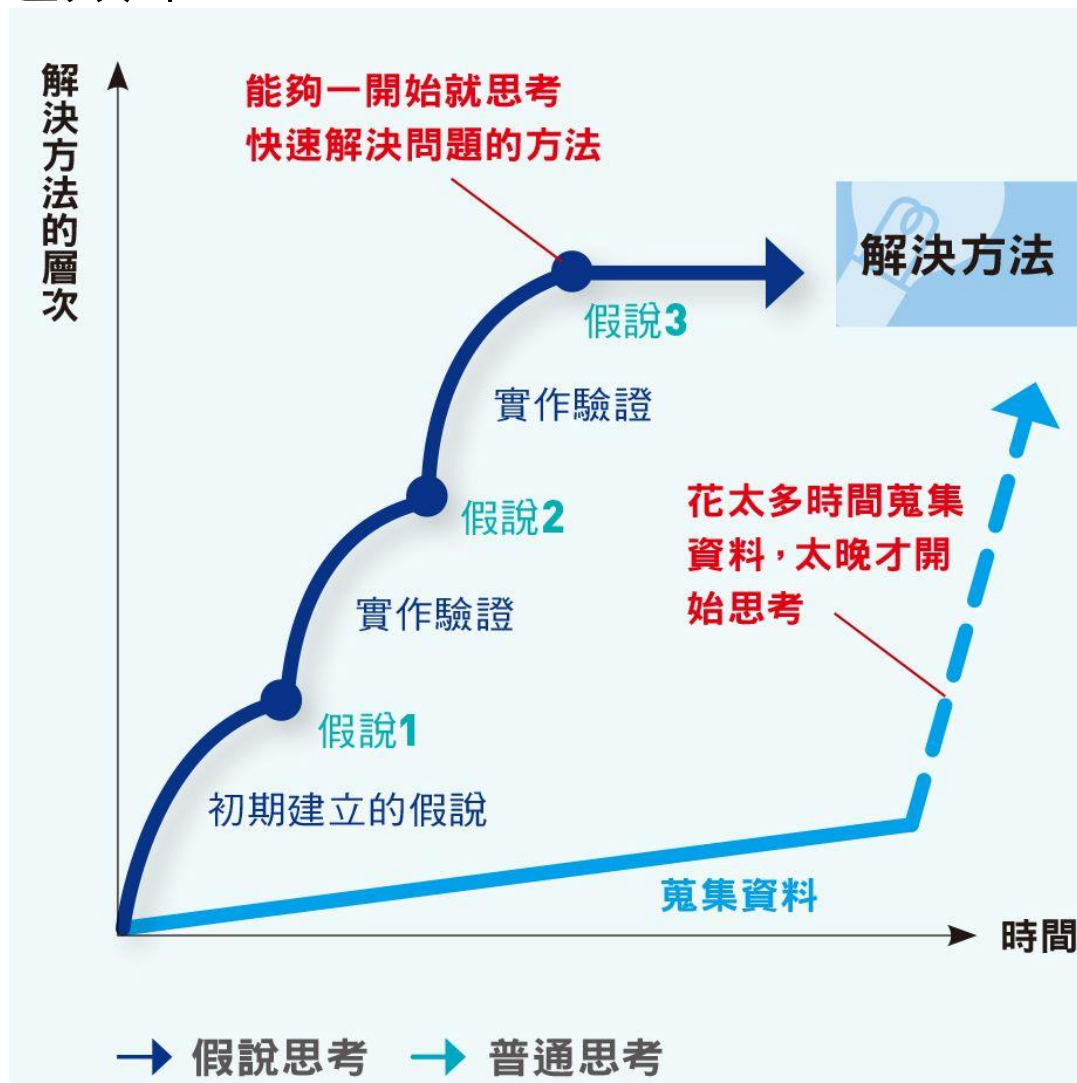
如何增進溝通效率



如何增進溝通效率



如何增進溝通效率



如何增進溝通效率



01 MECE (彼此獨立、互無遺漏)

子集合之間的關係像是拼圖一樣，沒有重複也無遺漏。



02 重複

子集合之間有所重疊成。



03 遺漏

子集合之間有所漏失。

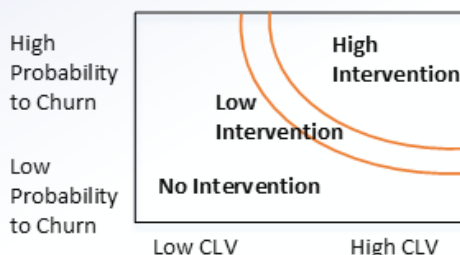
美國某電信業-分析客戶流失

Figure 1: Considering probabilities and risk tolerance



Business assesses the tolerance of risk with regard to churn

Figure 2: Decision model for Churn management



Decision model is based on customer lifetime value

Figure 3: Multiple actions based on customer segments

SEGMENT	ACTION
Greater than 50% probability to purchase and top 3 decile customer	Email and call
Greater than 25% probability to purchase and top 5 decile customer	Email
Unlikely to purchase and/or low decile customer	No promotion

Customer Segmentation (example of a policy)

<https://docs.microsoft.com/zh-tw/azure/machine-learning/studio/azure-ml-customer-churn-scenario>

<https://gallery.azure.ai/Experiment/31c19425ee874f628c847f7e2d93e383>

美國某電信業-分析客戶流失

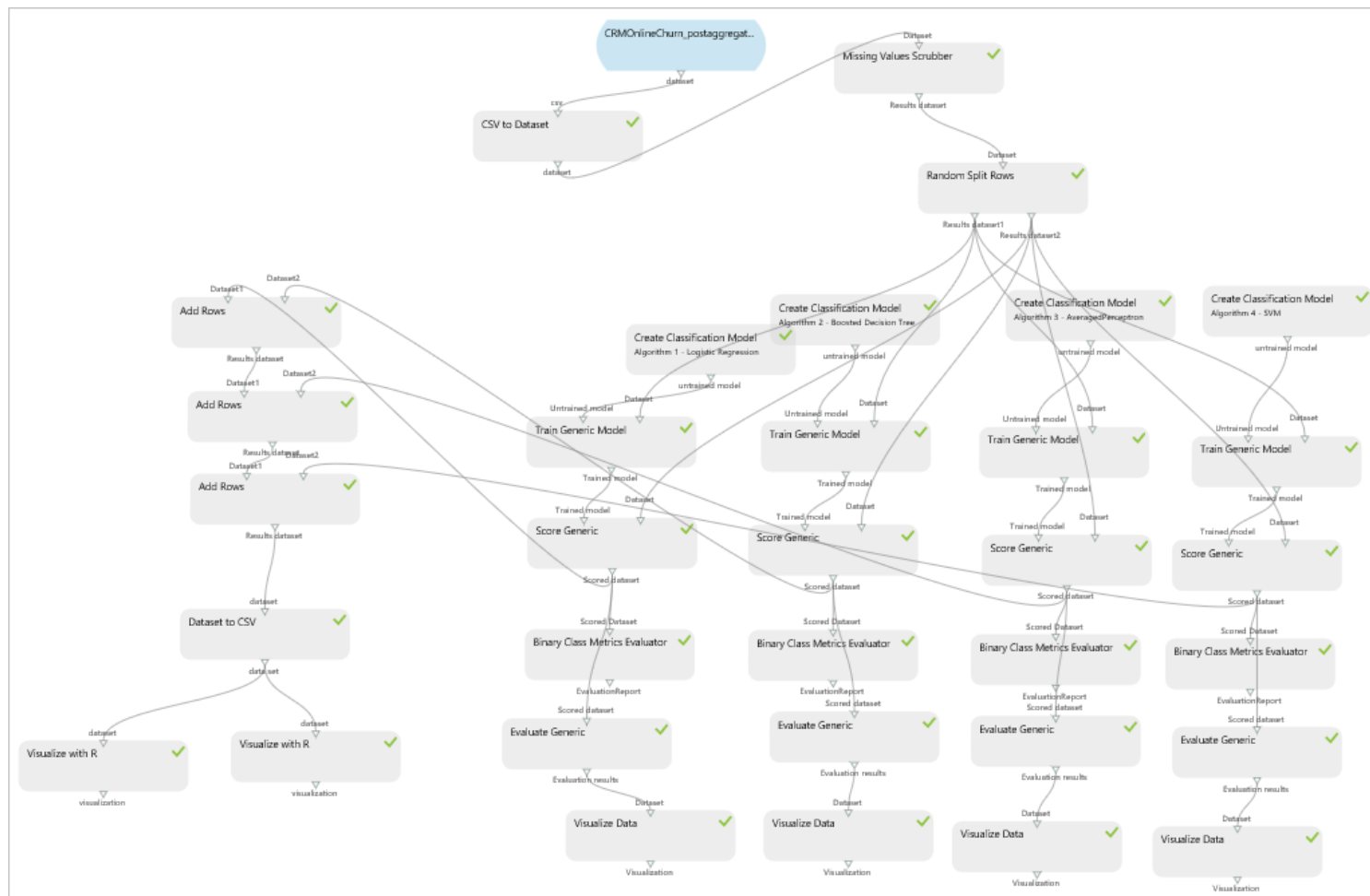
Churn Binary	GroupCustomerDBID	Mean(Ac	Impute_f	IMP_Mean	Mean(Activ	Impute_	IMP_	Mean(Cases)	Imp	IMP_Mean(I	Mean(Co	Impute_
1	Commerce.3766384	2.833213	0	2.833213	7.628518	0	7.6285	0	0	0	4.430817	0
1	Commerce.3648614	2.197225	0	2.197225	5.147494	0	5.1475	0	0	0	3.401197	0
1	Commerce.1701750	4.682131	0	4.682131	7.451242	0	7.4512	2.197225	0	2.197225	5.056246	0
1	Commerce.1451565	4.787492	0	4.787492	7.960803	0	7.9608	0	0	0	5.398163	0
1	Commerce.2115489	3.433987	0	3.433987	3.7612	0	3.7612	1.791759	0	1.791759	3.091042	0
1	Commerce.6205107	3.988984	0	3.988984	7.467873	0	7.4679	0.693147	0	0.693147	4.89784	0
1	Commerce.7125701	4.356709	0	4.356709	4.077537	0	4.0775	0	0	0	6.603944	0
1	Commerce.2808747	0.693147	0	0.693147	7.427045	0	7.427	0	0	0	7.029973	0
0	Commerce.3213575	7.17012	0	7.17012	5.590987	0	5.591	0	0	0	7.284135	0

<https://docs.microsoft.com/zh-tw/azure/machine-learning/studio/azure-ml-customer-churn-scenario>

<https://gallery.azure.ai/Experiment/31c19425ee874f628c847f7e2d93e383>

美國某電信業-分析客戶流失

使用Azure ML 建置客戶流失預測模型

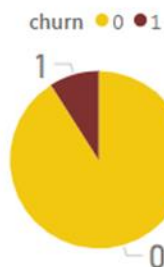
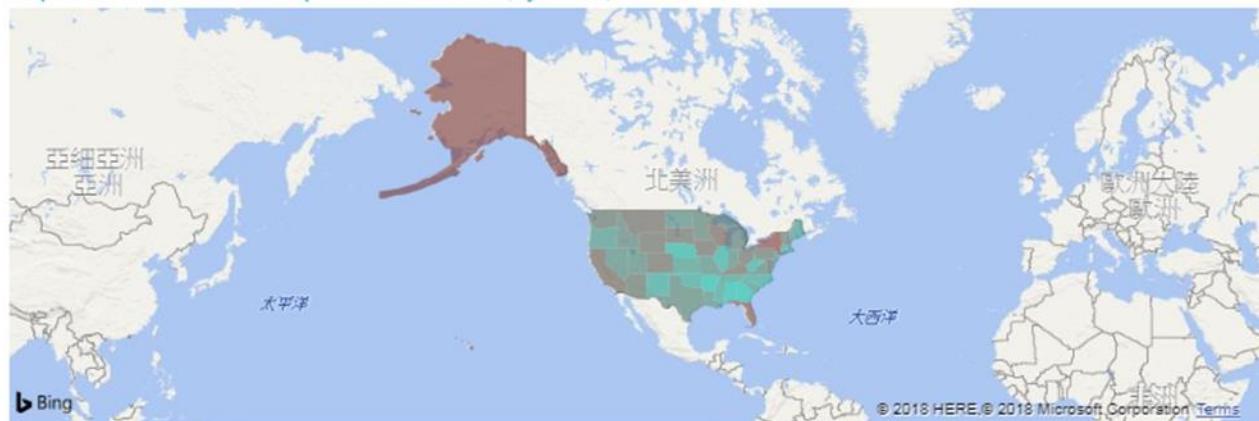


<https://docs.microsoft.com/zh-tw/azure/machine-learning/studio/azure-ml-customer-churn-scenario>

<https://gallery.azure.ai/Experiment/31c19425ee874f628c847f7e2d93e383>

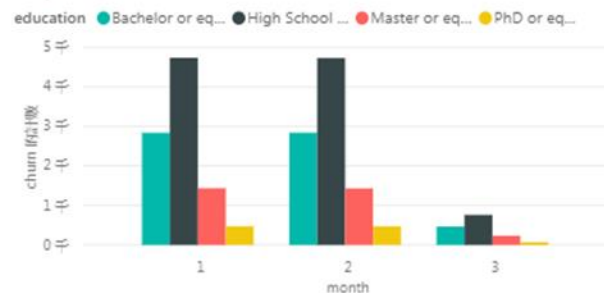
使用Power BI監控客戶流失行為

Impact of Customer Complaints on Churn (By State)

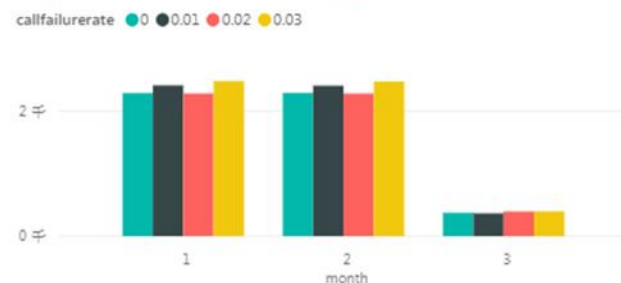


Click the chart to analyze what impacts customer churn

Impact of Education Level on Churn



Impact of Call Failure Rate (%) on Churn



<https://app.powerbi.com/view?r=eyJrljoiMzVhODJmMjctYjU0YS00NTVhLWlZTTEtZjVkYTZhYyBjMmJiliwidCI6IjU3NGMzZTU2LTQ5MjQtNDAwNC1hZDFhLWQ4NDI3ZTkYjl0MSIsImMiOiZ9>

Tableau– 5 ways Tableau look at Tableau's own sales data

- Sales performance: How are we doing this quarter?



Tableau– 5 ways Tableau look at Tableau's own sales data

- Segmentation: Which segments bring in the most revenue?



<https://www.tableau.com/about/blog/2015/9/top-metrics-every-sales-executive-cant-live-without-43746#rule7jCQO2MoPWxw.99>

LinkedIn-- empowers thousands of salespeople with Tableau, reducing customer churn

90% of sales team
empowered by Tableau
Server

Centralize petabytes of
data from disparate
systems

More proactive churn &
performance analysis

領域知識 + X

領域知識: e.g. 消費品行銷、影音OTT串流、旅遊業、
電商平台、科技製造、金融商品...

統計

程式語言
(R/Python)

資料庫語言
SQL

建模軟體
(SAS/ SPSS
Modeler/ Azure ML)

溝通技巧

分析問題技能

機器學習/
深度學習

資料視覺化
(R/ PowerBI/
Tableau)

ETL &
Infrastructure...

Contact Me



Angie Hsu 徐紹婷

<https://www.linkedin.com/in/angie-hsu-shaoting/>