數據分析世界裡的‘Hello World’: 手把手教你數據分析流程

**課前閱讀:**

1. 您能找出幾種KDD?   
   KDD → Knowledge Discovery in DataBase 資料庫知識探索

從資料庫中，探索出有用知識的程序。

1. 請觀看:

<https://www.youtube.com/watch?v=39TcJmr60wA>

資料分析師、資料工程師、資料科學家，哪個適合我？

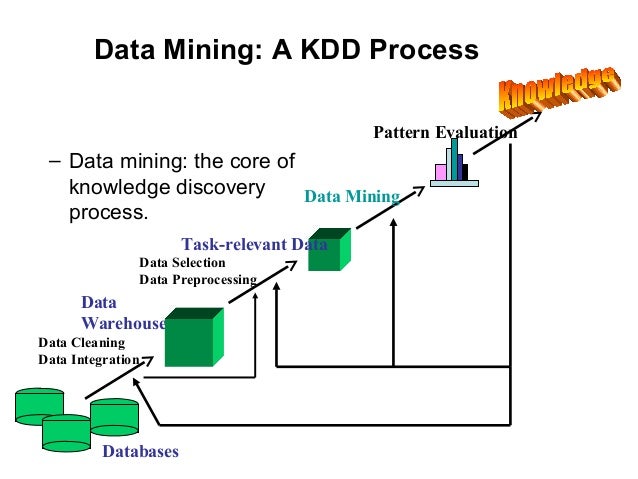
* 資料科學家：詮釋這個資料給管理階層，影響決策，說故事的能力，告訴大家這個資料背後的故事。需要具備做研究的經驗跟能力。
* 資料分析師：熟悉數學統計（機器學習）及演算法，甚至視覺化能力，讓人一目瞭然資料結果。
* 資料工程師：幫助資料分析師做前段的工作，例如整理格式，也就是ETL（Extract-Transform-Load 擷取、轉換、載入，變成可分析的資料）；具備IT資訊工程能力，包含雲端、資料庫管理、分散式運算的方法。

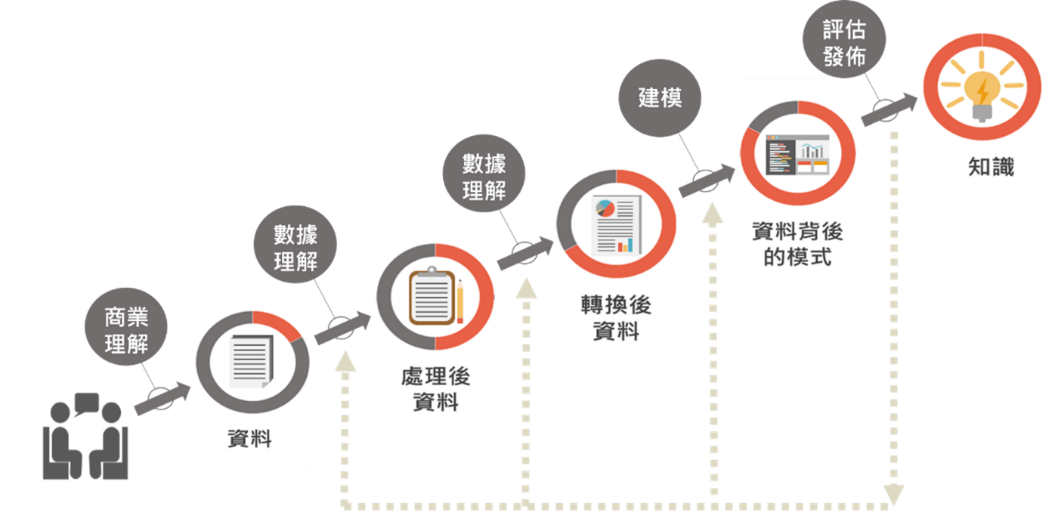
1. 請觀看厚數據: <https://www.ted.com/talks/tricia_wang_the_human_insights_missing_from_big_data>

大數據能在一定程度上洞悉問題，並最大程度發揮機器智能，而厚數據能幫我們找到那缺失的背景資訊，能讓大數據便於使用，並最大程度發揮人類智能。

**課程活動/作業:**

**活動1: 請完成以下工作單程序**

****

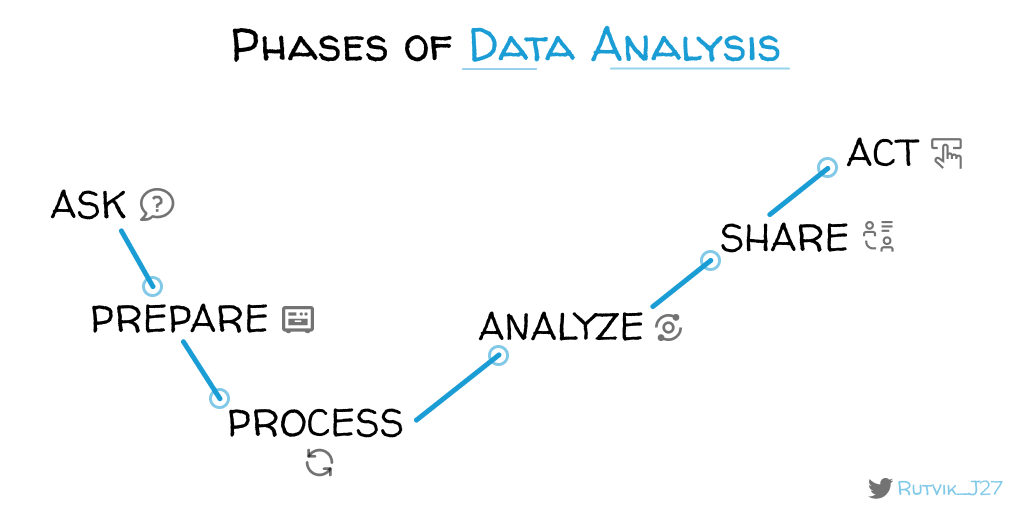


* 先做清理，再做轉換。

請將上述8個步驟排序進入KDD程序~

1. **訂定目標（identifying the goal）: 商業理解**
2. **建立目標資料集（creating a target data set）**
3. **資料清理與前置處理（data cleaning and preprocessing）→〈數據理解〉處理後資料(Process)**
4. **資料轉換（data transformation）→〈數據理解〉轉換後資料**
5. **選擇資料探勘方法（Choosing the data mining method）→〈建模〉資料背後的模式**
6. **資料探勘（data mining）→〈建模〉資料背後的模式  
   → 讓最終使用者了解根據前面各程序步驟所獲得的資料探勘結果。**
7. **解釋探勘模式（interpreting mined patterns）→〈建模〉資料背後的模式  
   → 對最終選定的資料探勘模式進行解釋。**
8. **鞏固發現的知識（acting on the discovered knowledge）  
   → 運用KDD最終發現的知識結果並採取行動。同時，也要檢視該知識結果與過去的觀點是否一致。**

**Ask→Prepare→Process→Analyze→Share→Act**



課程裡的疑問:

|  |  |
| --- | --- |
| 問題1 |  |
| 回答1 |  |
| 問題2 |  |
| 回答2 |  |

本堂課程裡我學到了:

延伸問題/學習/應用