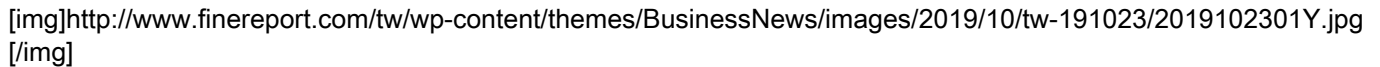


frilly / October 25, 2019 08:10PM

[【推薦收藏！】數據分析師必備的10大技能，大多數人只知道一半](#)



昨天有位剛入行資料分析的朋友跟我吐槽，自己入門到現在只會用Excel和做分析圖表，但是感覺越做越沒有價值，資料分析似乎就是業務資料的分析員，不知道該如何提升自己。

這是因為他沒有完全把<http://www.finereport.com/tw/>大數據分析[\[url\]](#)的價值挖掘出來，資料分析是為了通過對資料現象的查看來完成對產品、行銷策略、運營策略的優化，不僅是對業務，更重要的是要掌握資料分析的各種技能，從能力增長上突破職業的天花板。根據我多年總結的經驗，一個合格的、高級的數據分析師必須要掌握以下10種技能：

[s]

統計分析：大數定律、抽樣推測規律、秩和檢驗、回歸、預測

視覺化輔助工具：excel、BI工具、python

大數據處理框架：Hadoop、storm、spark

資料庫：SQL、MySql、DB

資料倉庫：SSIS、SSAS

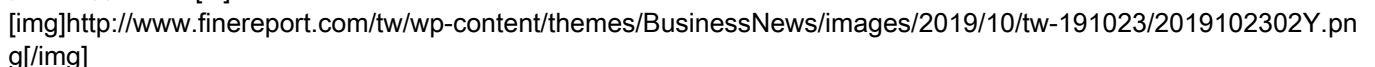
資料採擷工具：Matlab、R語言、python

AI人工智慧：機器學習

挖掘演算法：資料結構、一致性

程式設計語言：Java、python

撰寫報告的能力[/s]



[size=large]1、統計分析[/size]

眾所周知，統計學是資料分析的基石。學了統計學，你會發現很多時候的分析並不那麼準確，比如很多人都喜歡用平均數去分析一個事物的結果，但是這往往是粗糙的。而統計學可以幫助我們以更科學的角度看待資料，逐步接近這個資料背後的“真相”。

大部分的資料分析，都會用到統計方面的以下知識，可以重點學習：

基本的統計量：均值、中位數、眾數、方差、標準差、百分位數等

概率分佈：幾何分佈、二項分佈、泊松分佈、正態分佈等

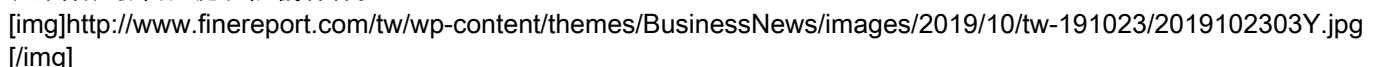
總體和樣本：瞭解基本概念，抽樣的概念

置信區間與假設檢驗：如何進行驗證分析

相關性與回歸分析：一般資料分析的基本模型

瞭解統計學的原理之後，你不一定能夠通過工具實現，那麼你需要去對應的找網上找相關的實現方法，也可以看書。先推薦一本非常簡單的：吳喜之-《統計學·從資料到結論》，也可以看《商務與經濟統計》，結合業務能更容易理解。

另外，如果想要更進一步，請掌握一些主流演算法的原理，比如線性回歸、邏輯回歸、決策樹、神經網路、關聯分析、聚類、協同過濾、隨機森林。



再深入一點，還可以掌握文本分析、深度學習、圖像識別等相關的演算法。關於這些演算法，不僅需要瞭解其原理，你最好可以流暢地闡述出來，還需要你知曉其在各行業的一些應用場景。如果現階段不是工作剛需，可不作為重點。

[size=large]2、[\[url=http://www.finereport.com/tw/products/datavisualization\]](http://www.finereport.com/tw/products/datavisualization)資料視覺化[\[url\]](#)輔助工具[/size]

資料視覺化主要通過程式設計和非程式設計兩類工具實現，對於普通行業的資料分析師來說，可以不需要掌握程式設計類的視覺化工具，學習麻煩浪費效率，推薦掌握以下即可：

[size=medium]1、Excel[/size]

作為常用的資料處理和展現工具，資料分析師除了要熟練將資料用 Excel 中的圖表展現出來，還需要掌握為生成的圖表做一系列的格式設定的方法，如：系列格式美化、三維格式美化，以及坐標軸和格線設定等，圖表可以與函數或巨集等功能一起聯用，製作出類比圖表或帶有交互效果的高級圖表，比如在臺灣地圖上標注各市縣的人口分佈等，實現這些能得到更好地資料分析和查看效果。Excel 裡面自帶的資料分析功能，很大程度上能完成專業統計軟體 (R、SPSS、SAS、Matlab) 的資料分析工作，這其中包括描述性統計、相關係數、概率分佈、均值推斷、線性、非線性回歸、多元回歸分析、時間序列等內容。熟悉使用 Excel 的各項功能對一名優秀的資料分析師來說非常重要。

[size=medium]2、BI工具[/size]

近幾年冒出來的BI之秀，如powerBI、qilkview都強調視覺化，一改傳統BI工具SAP BO、IBM家的cognos (不過近幾年貌似都在研發雲BI)。這裡不談開源，還沒見到能成熟應用的BI。成熟的BI軟體工具如FineReport (國內) 和 Tableau (國外)，都很推薦。

[url=http://www.finereport.com/tw/knowledge/acquire/tableauvsfinereport.html]Tableau VS FineReport?

[/url]

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2019/10/tw-191023/2019102304Y.gif[/img]

FineReport作為一款通用的報表及資料視覺化工具，兩大核心功能是填報和資料展示，但我覺得比較驚豔的一點是，它內置了大量的圖表和視覺化動效，視覺化很豐富，完全沒有印象中做報表那種古板的風格。多以它能做出格式各樣的dashboard、甚至是視覺化大屏，一點不虛。

感觸最深的是之前工作拿FR開發報表很省力，10張門店報表以往做10張Excel的，在他裡面就是一個參數查詢，然後批量匯出，用一個範本。

另外這款工具除了可以做視覺化報表，同時也提供大屏的服務。通過佈局、色彩、綁定資料等環節完成大屏戰情室的製作，擁有很多自助開發的視覺化外掛程式庫。

因為後端通常連結業務系統資料，所以可以即時連接業務資料，做企業的一些經營資料展示。比如展覽中心、BOSS儀錶板，還有城市交通管控中心、交易大廳等。

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2019/10/tw-191023/2019102306Y.gif[/img]

[size=medium]3、python[/size]

學過Python資料分析的朋友都知道，在視覺化的工具中，有很多優秀的三方庫，比如matplotlib，seaborn，plotly，Boken，pyecharts等等，這些視覺化庫都有自己的特點，在實際應用中也廣為大家使用。如果你不知道資料分析該學什麼工具，就直接學python吧，萬能語言學了不虧。

學習資源推薦：

[url=https://www.facebook.com/twfinereport/]Python Weekly[/url]：每週更新，包括Python相關的文章、教程、演講、書籍、專案、工作等。

Python challenge：用Python程式設計技能解決33個謎題，闖關。

[url=https://www.python.org/]Python官方文檔[/url]：官網，去找資料吧！

[url=https://pythontips.com/2013/07/31/10-python-blogs-worth-following/]10 python blogs worth following[/url]：

關於python你值得關注的10個博客

[size=large]3、大數據處理框架[/size]

如果你想脫離普通業務的束縛，做一名大數據分析師，首先就要瞭解大數據框架的基礎。

大數據處理框架負責對大數據系統中的資料進行計算，資料包括從持久存儲中讀取的資料或通過訊息佇列等方式接入到系統中的資料，而計算則是從資料中提取資訊的過程。

我們按照對所處理的資料形式和得到結果的時效性進行分類，分為批次處理系統、流處理系統和混合式系統。典型的批次處理系統就是Apache Hadoop；典型的流處理系統有Apache Storm，Apache Samza；混合處理系統比如Apache Spark，Apache Flink。

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2019/10/tw-191023/2019102308Y.jpg[/img]

[size=large]4、資料庫[/size]

資料分析是分等級的，有只負責清洗資料的，比較少，工作也比較簡單；還有就是負責建模的，掌握常用的十多個機器學習演算法就能是二流的了，要做到一流的就要熟練掌握各個演算法的本質了，也就是要掌握資料庫的基礎。sql在資料庫裡是核心技術，在資料分析學習時一定要重視這些內容，主要以MySQL為主，MySQL就是互聯網行業的通用標準。

目前使用較多的資料庫有MySQL、SQL Server 和 Oracle，資料分析師必須掌握的常用語句和函數有如下幾種。

- (1) 合計和標量函數：Count()、Max()、Sum()、Upper()、lower()、Round()等。
 - (2) distinct——distinct 關鍵字可以過濾重複的資料記錄。
 - (3) Top——結合 select 語句，Top 函數可以查詢頭幾條和末幾條的資料記錄（僅限 SQL Server，在其他資料庫，可用 limit 語句、rownum 列等方式實現相似的目的）。
 - (4) Order By——結合 select語句，Order By 可以讓查詢結果按某個欄位正序和逆序排列。
 - (5) Group By——Group By 子句可以對查詢的結果集按指定欄位分組。
 - (6) Group By & Having 子句——Having 語句基於 Group By，定義分組條件。
 - (7) Inner Join，Left Outer Join，Right Outer Join and Full outer Join——多表的列關聯，即通過 Join 可以將不同物理表中的資料列根據一定的關聯條件合併成一個結果集。
 - (8) Union 合併查詢：Union/Union ALL 查詢可以把多張表的資料行合併起來，Union 在合併時重複的資料僅保留一行，而 Union ALL 則是直接合併，不會處理重複行。
- 在大數據時代，有很多查詢工具可供選擇。Hive 和 SQL 是目前比較主流的工具。Hive 是基於 Hadoop 的一個數據倉庫工具，可以將結構化的資料檔案映射為一張資料庫表，並提供完整的 SQL 查詢功能，可以將 SQL 語句轉換為 MapReduce 任務進行運行。Hive 和 SQL 是非常相似的，主要的區別就是 Hive 缺少更新和刪除功能。如果你可以熟練使用 SQL，就可以平穩過渡到 Hive。
- 另外，一定要注意兩者在結構和語法上的差異。

當然，如果你想要快速掌握資料庫的知識，一定要進行系統化的學習以及大量的練習，在網上尋找一些資料庫的練習題，先從簡單的題開始，循序漸進，這樣才能夠慢慢的深入資料庫的核心知識。

[size=large]5、數據倉庫/BI商業智慧[/size]

在進行資料分析的時候，我們總會遇到一些名詞，比如資料倉庫。數據倉庫是資料分析中一個比較重要的東西，數據倉庫是一個面向主題的、內建的、相對穩定的、反應歷史變化的資料集合。

資料分析中的工作最重要的就是資料處理工作，根據我做資料分析的經驗，在整個資料分析流程中，用於資料處理的時間往往要佔據70%以上，而數據倉庫具有內建、穩定、高品質等特點，基於數據倉庫為資料分析提供資料，往往能夠更加保證資料品質和資料完整性。

[img]http://www.finereport.com/tw/wp-content/themes/BusinessNews/images/2019/10/tw-191023/2019102309Y.jpg[/img]

[size=large]6、資料採擷工具[/size]

[/size]

在做資料分析時，資料採擷軟體是其中必不可少的工具之一。它是大多數商業智慧計畫中的核心應用程式，資料採擷軟體同樣也能夠從大量資料中提取洞察力。

直接說需要學習的語言：MATLAB、Python、R。

1、MATLAB

雖然偏學術性，但是好上手，上手以後就可以跑一些演算法，提高一些信心和學習的樂趣。教材看官方手冊的Primer，然後就開始寫腳本和函數，如果有看不懂的直接百度、google或者help。

2、Python和R

這兩個放在一起，是因為網上關於這兩個的爭論太多了。我的順序是首先學python，其次再是R。首先python，先看《Head First Python》，簡單易懂，然後是《利用Python進行資料分析》和《機器學習實戰》。第一本書主要是利用Python做資料採擷的，基本提到Python學習都會推薦這本。第二本是理解機器學習的佳作，書中用到的語言就是Python，一邊學語言，一邊理解機器學習。

[size=large]7、AI人工智慧[/size]

嚴格意義上，人工智慧與資料分析有著明顯的界限，不屬於同一領域，因此這一條是針對大資料分析科學家來說的，當你的分析能力還比較低時，可以略過不看此章。

機器學習、人工智慧涵蓋的知識層面太廣太深，所以會建議採取 problem-based learning 的學習方式，先選定問題，然後找到資源來解問題，再更深入的去瞭解。這樣在解決問題過程中，遇到的名詞與盲點知識會積累的愈多。

[size=large]8、数据挖掘算法[/size]

很多人认为数据挖掘需要掌握复杂高深的算法，需要掌握技术开发，才能把数据挖掘分析做好，实际上并非这样，其实算法并不难，只需要结合实际业务背景、以解决问题为导向就简单很多了，主要包括分类算法，聚类算法，关联规则三大类，这是学习数据挖掘必须要掌握的算法基础，这三类基本上涵盖了目前商业市场对算法的所有需求。而这三类里又包含许多经典算法。

我在这里推荐三本比较经典的书吧：

Pattern Recognition and Machine Learning：经典中的经典。

The Elements of Statistical

Learning：算是入门ML的功夫秘籍，经典，书中配有R包，非常赞！可以参照着代码学习算法。

Machine Learning：作者Kevin Murrphy教授是机器学习领域中年少有为的代表。这书是他的集大成之作，写完之后，就去Google了，产学研结合，没有比这个更好的了。

[size=large]9、編程語言[/size]

比如python、r語言、java等等，你該使用哪種語言用於數據分析？恐怕這還得“視情況而定”。

如果你对晦涩的统计运算进行繁重的数据分析工作，那么你不青睐R才怪。如果你跨GPU进行NLP或密集的神经网络处理，那么Python是很好的选择。如果想要一种加固的、面向生产环境的数据流解决方案，又拥有所有重要的操作工具，Java或Scala绝对是出色的选择。

以 R 语言为例，R 编程语言在数据分析与机器学习领域已经成为一款重要的工具。R 作为脚本语言凭借其良好的互动性和丰富的扩展包资源可以方便地解决大部分数据处理、变换、统计分析、可视化的问题，并可以重现所有的细节。R 的优势在于有包罗万象的统计函数可以调用，特别是在时间序列分析方面（在游戏行业也有很好的应用），无论是经典还是前沿的方法都有相应的包可以直接使用。因此，掌握 R 语言可以提高整体的生产力。

然而，要成为一名优秀的数据分析师，仅学会使用一门语言远远不够，还需要修改数据挖掘语言的程序包或模型，因为现有的程序包或模型有局限性，在前期数据处理上还是不够自由，如异常值的处理、变量处理等，而自己写代码编程也可以根据自己的需求进行编写，实现更多的个性化需求。

[size=large]10、撰写报告[/size]

撰写报告的能力对成为一名优秀的分析师来说也非常重要。

資料分析報告是對整個資料分析過程的一個總結與呈現。通過報告，把資料分析的起因、過程、結果及建議完整地呈現出來，供決策者參考。一份好的資料分析報告需要有如下3點要求：

（1）好的分析框架：首先需要有一個好的分析框架，並且圖文並茂、層次明晰，能夠讓閱讀者一目了然。結構清晰、主次分明可以使閱讀者正確理解報告內容；圖文並茂，可以令資料更加生動活潑，提高視覺衝擊力，有助於閱讀者更形象、直觀地看清楚問題和結論，從而產生思考。

（2）明確的結論：沒有明確結論的分析稱不上分析，同時也失去了報告的意義，因為我們最初就是為尋找或者求證一個結論才進行分析的，所以千萬不要舍本求末。

（3）建議或解決方案：作為決策者，需要的不僅僅是找出問題，更重要的是建議或解決方法，以便他們在決策時作參考。所以，資料分析師不僅需要掌握資料分析方法，而且還要瞭解和熟悉業務，這樣才能根據發現的業務問題，提出具有可行性的建議或解決方案。

即便有嚴謹的分析思路和有價值的資料資料，如果不能將其寫成報告，或者寫的報告未能準確清楚地表達出資料中隱含的規律，那資料的價值將大打折扣。一份好的分析報告，資料資料是功底，報告的框架是支柱，報告的格式是軍裝，獨特見解是亮點，預測方法是刀槍，正確的判斷是見證。在撰寫報告時，深入地思考，深入分析，邏輯嚴謹，結論有說服力，能提前預測資料趨勢，能從問題中引申出解決方案，提出有指導意義的分析建議，這些都是一名優秀的分析师所體現的特質。

除了以上的硬實力，資料敏感力、邏輯思維能力、歸納能力、批判性思維能力、交流溝通能力、責任力這些軟性的技能也是優秀分析師必須具備的素質。另外，如果分析師能站在更高的角度思考問題，有管理者的思維，那麼就能在眾多分析師中能脫穎而出。

以上有些素質是我們在入職場之前就具備的，而有些則需要進入行業環境後逐步積累和建立。成為優秀的資料分析師需要具備過硬的業務素養和技術能力，這絕非一朝一夕之功，需要在實踐中不斷成長和昇華。一個優秀的資料分析師應該以資料價值為導向，放眼全域、立足業務、善於溝通，認真對待每一次的資料分析工作，在工作中快速成長。祝願各位！

>>>

驚喜來了！！

我為大家專門製作了數據分析師必備教程合集電子書，裡面囊括這兩年來為數據人提供的經驗總結和科普文，共40篇，現在妳只需私訊我（暗號：data）即可獲取這一份超級超級豐富的禮物，傾心準備的第一版，希望你們喜歡~

[url=http://www.finereport.com/tw/products/trial]FineReport10.0免費下載[/url]

獲得帆軟最新動態：數據分析，報表實例，專業的人都在這裡！加入[url=https://www.facebook.com/twfinereport/]FineReport臉書粉絲團[/url]！

Edited 1 time(s). Last edit at 10/27/2019 08:35AM by frlily.
