2016-05-23 傅一平 与数据同行



阅读本文前,请您先点击本文标题下面的蓝色字体"与数据同行"再点击"关注",这样您就可以分享一 个大数据从业者的真实数据生活,独家数据观点!



大数据时代到来,如何从数据中提取、挖掘对业务发展有价值的、潜在的知识,为决策层的提供有力依据, 为产品或服务发展方向起到积极作用,有力推动企业管理的精益化,对于每个企业都意义重大,而这些工 作,大多需要数据分析师才能完成,但如何才能成为一个合格的数据分析师呢? 我这里提四个方面,如果你是一个新手,想从事数据分析师这个职业,那么,你可以看看,当然如果你是个

分析老鸟,在苦恼如何更进一步,也可以看下,可能于你也有益哦,数据分析师学无止境,总在痛并快乐

Part 1 数据分析师的基本素质 数据分析师最需要的基本素质是什么,很多人会回答,不是要统计学知识吗,不是要会一些工具吗,说得

在技巧之外,即业务能力、思考能力及沟通能力三样,这些能力很多人天生就有,而很多人需要艰苦的训练 获得,甚至训练也不可得。 业务能力

不要听什么数据分析速成的教程,任何有经验的数据分析师对此都会嗤之以鼻,要成为数据分析师,没个3年

对,也不对,统计学等只是数据分析的术,而我更强调的却是道,做好数据分析与写诗文一样,真正的关键

下不来,当然具体周期依赖于公司业务的复杂程度。只要真的在实践领域从事过数据分析工作,就会明白所 有分析的重中之重都是业务知识本身。而业务知识的学习和掌握,需要的积累之深,培养一个业务专家,需

要的周期之长,都远远超过后面所说的那些基本技能,成为业务专家实属不易,数据分析师其实是之于业务 专家之上的更深层次的思考和总结,否则,谁指导谁都是个问题。业务学习的方式很多,比如将以前的分析 报告和取数案例都拿过来研究一下,不懂就问,总是一个渐进的过程,但需要时间和行业的沉淀。数据分析 师最需要不断提升的能力就是行业和业务知识,没有之一。 思考能力

数据总是在那里,它不会说话,你不仅要基于业务能力理解它,还要学会推演和分析,从中发现规律,迅速

定位某个商业问题的关键属性和决定因素,形成自己独创性的见解,所谓心思缜密,滴水不漏,没有思考逻 辑没有数据分析。而要形成独特的见解,则来自于个人不断的学习和思考,这里的学习更多的强调是跨领域

和专业,思考则更多的强调养成思考的习惯。

跨领域的的知识会给人带来不同的思维方式和看问题的角度,每学习一个新的领域,相当于打开了一扇全新 看世界的门,很多企业经常花重金请咨询公司帮忙做一些分析报告(当前咨询分析不再像以前吃香,不是企 业不需要,而是他们也需要与时俱进),是有其原因的,企业往往看重咨询公司广阔的分析思路和全局的视 野,比如经常提的最佳实践。在各个跨领域专业中,经济学、心理学、统计学等对于数据分析的辅助又最

另一个思考能力是养成思考的习惯,所谓"学而不思则罔,思而不学则殆"。思考本身是一种实践,它可以 将你的知识更加系统化和深入化,数据分析一定程度上是用来验证思路和启发灵感的,"数据分析"从来不 是"数据分析"本身,而是以"数据分析"为手段和表象,对业务的深刻理解、思考和判断。 沟通能力

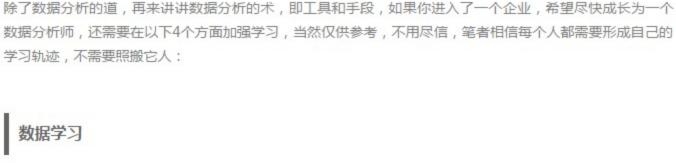
数据分析贯穿BIT、数据、技术、业务整个链条,数据分析师将BIT最终转化成决策者理解的语言,跨越的流 程很长,你需要面对不同的岗位,碰到不同的角色,采用不同的语言,表达你的要求和获得你需要的东西, 成为数据和业务的桥梁,没有足够的沟通能力很难。所谓上得了台面,下得了厅堂。同时,大脑很多时候是 非理性的,但如果你容易听取他人的意见,特别是智者的意见,则可以帮你找到另一条出路,你犯错的概率

就会降低,相应的,你的分析就更有力量和说服力 从这个角度讲,企业的管理者实际是最强的数据分析师,这个毋容置疑哦。数据分析师道的养成,绝非一日

数据学习

之功。 Part 2

数据分析师的养成课程



的业务含义,理解的越透彻,你的分析潜力就越大。更有甚者,笔者还建议你去理解源系统,从业务实现流 程角度出发去理解对应数据的含义,因为有时候,简单的业务描述在数据上的表达却是非常复杂的,业务语 言与数据语言很多时候是1对多关系,打个比方,你看到业务系统上某个菜单的功能,对应到系统中的数据是 怎样的,你能还原出来吗?

当然,大多数数据分析师可能不需要进行系统数据学习,反正实践中慢慢熟悉好了,但自顶向下的数据学习

远只能在已知的数据里转圈圈,当然,可能也够了,但我这里要说得是做得最好。

业务学习有一个毛病,比如你看案例,往往接触到的数据或使用的数据是局部的,因此,你的视野会受局 限,在大多数公司里,很多数据分析师其实缺乏全局的数据视野,因为他不知道到底有多少数据,因此,永

数据分析师应该主动的向IT部门拿到最全的数据字典,对于数据字典进行持续的学习,了解每张表甚至字段

技术学习 有几个层面的东西要学,依赖于实际的场景和你希望达到的阶段: 首先,你要学会从数据库或者其它源头获取数据,很多数据分析师仍然依赖于IT人员获取数据,但大数据时

代,真的有必要自己动手了,因为依赖他人效率太低了,起码你要会SQL,SQL甚至基本上是为统计取数而生 的方便工具,图形化的透视方式也远远没有SQL的表达能力强,这是基本功,其实,SQL就能解决大多数统

计取数问题。 其次,你要会一些数据分析工具,EXCEL是最基本的,其实大多数数据分析基于EXCEL应付已经绰绰有余了,

作用,难度容易。

EXCEL的图形表达能力也已经够强。

最后,如果你希望更深一层,那就学习R语言、PYTHON、SPSS,SAS等,他们提供了更强大的挖掘能力, 可以帮你把统计学的数据挖掘精髓发挥的淋漓尽致。

终于讲到大家都很关心的统计学知识了,推荐一些书吧:

方式可以让你有一个更好的基础和更全局的数据视野。

当然,如果有可能,你也要熟悉一下你所在企业的数据仓库或大数据平台,懂得一些基本的操作,对于你提 升分析的自由度和灵活性也大有好处,比如自己搞个脚本定时跑数据,打造个人的数据集市,现在数据分析 的概念也越来越大,很多公司把对于大数据平台的数据处理能力也纳入到数据分析师的技能范畴。 以上层层递推,其实数据分析师每在IT上前进一步,带来的效益是几何级的,比如你懂Hadoop,那么,你就 可能离大数据更近一点。

统计理论

《深入浅出统计学》:要了解常用数理统计模型(描述统计指标、聚类、决策树、贝叶斯分类、回归等), 重点放在学习模型的工作原理、输入内容和输出内容,至于具体的数学推导,学不会可暂放一边,难度容 易。

《数据挖掘导论》:最近几年数据挖掘教材中比较好的一本书,被美国诸多大学的数据挖掘课作为教材,笔

《深入浅出数据分析》:讲了数据分析到底是干什么的?数据分析都包含什么内容?对新人们还是有一定的

者也最近买的,很好,很多概念解释的比较清楚,难度中上。 《统计学习方法》:李航老师的扛鼎之作,难度难。

多了,大家自己搜索。

这些都是很实用的书籍,但结合了实践学更好,对于特定的业务场景,就找对应的书看吧,网上推荐的也很

做重点突出,如何图文并茂,PPT的写作决非一日之功。

下面是我以前研究的一个新人学习计划表,供参考。

L表开取数值力

2.9KF开发培养

上加州银五位

LIBRIGHT

上部队在效治通

水·ZMTI 数据透视实际机

上售选开发和建模方法

1. 熟悉公司。加门及科安情况

2. 熱意教斯企车斯坦知识

上共选级。有描写规范

5. 数据仓库基础特别

3.000 美国电影中国

4.常用配置培训

5.均值根据证明

6. 思维导用导出

7. 北州八世北省

12. 浙江移动应规始两 is. Misan 14. 如何在他 机交换铁路

L 中国移动性与WOMSS被超影利用面 9. 中国森功图江公司大量直接哲学真结训 10、中国移动浙江公司联系统全型数据字典 11、浙江移动[[宋]]和系统被构刻间

当然也需要定期对学习成果进行检查

1、做过多少个分析或项目?

3、做过多少种类型的模型?

4、基于模型做过多少次完整的marketing闭环?

5、到底你的分析为企业带来了多大的实际价值?

以上五个问题,足以秒杀大多的忽悠和菜鸟了。

■ 艰难的抉择,阿里"小前台、大中台"的解读

■ 看上去很美,谈谈阿里云的大数据平台「数加」

■ 普及、开放与平台:大数据价值运营之路(上) ■ 普及、开放与平台:大数据价值运营之路(中) ■ 普及、开放与平台:大数据价值运营之路(下) ■ 六把武器?谈谈DT时代的大数据资产管理(上) ■ 六把武器?谈谈DT时代的大数据资产管理(下)

■ "数据化"与"差不多"先生,浅谈数据量化决策

■ 用心找书,大数据的思想书籍推荐

■ DPI大数据之战,运营商的艰难抉择

2、涉及多少业务场景?

历史文章

4. 热热双数

基础课程 上信息技术运动会

专业课程

工作与实践

表达能力

作为数据分析人员, PPT制作能力是极其重要的一项能力, 你总要通过某种形式表达出你的观点, 很核心的一 点是需要有严密的逻辑,甚至滴水不漏,可以通过思维导图来训练自己,但实际的格式表达却有点八股,一 般都是现状分析,原因剖析、分析结论和后续措施啥的,万变不离其宗,当然你需要花一点时间来了解如何

《极简统计学》: 对统计推断部分的阐述十分清晰, 适合非统计背景的人工阅读

《统计学:从数据到结论》:简明精要,统计概念和R可以一起学习

Part 3

数据分析师的学习计划

1. 学习沟通、总结、写作、深详等多础课程 1. 熟悉特殊赏销方法和混驳、辅助张与特确营辖工作

1. 提升取款, 开发和建模能力

工物科別市与企用表 工程等分別直向運程制制

RUSHERSUR.

2. 熟悉专题分析方还和高程。辅助参与专题分析工作

4. 熟悉物新拉拉力达和函程。研究参与收函控据工作

5. 姚升沟道。近结、号作、演讲等基础能力

#40] #53] #50] #51] #51] #50] #101]

BILLI

上就在导程指导下施工员员专通分析。相端营销成

2 独立负责专题分析。数据指挥、成构模型的活动

3.000 BRH 主要与解放扩张设计与模型

1. 了解解洗力集。据出参与解洗力发展写

数据的形成器

上加州西北州市州里

1. 专题分析系列课程结构 2. 数据的报名标准的基础 1. 粉病饮饮食利用取功证 1.提计分析推设 上海时以前、数据可是经济特别的工作 上海与部分水系统的、有规则是和存储的工程等工作



当然最重要的永远是实践,成为数据分析师最好的方式就是完成一个项目,如果有个老师就太好了,经过实

最后,再推荐一本科学技术思维的书,也是我最近看完的吴军的《数学之美》,其科普了部分模型,有利于

欢迎关注"与数据同行"公众号

如何访问?请关注"与数据同行" 微信公众号,点击历史文章菜单或者右上的按钮-查看历史消息

■ 从"男人比女人孝顺"和"百度医疗竟价"说起,大数据需要科学和正直的品格

践获得的知识才是你的,数据分析能力只能在实践的熔炉中提升、升华,永远问自己:

克服数学恐惧,其实从本原出发,数学和统计并不是那么的枯燥,传统教育真要反思哦。

■ 浙江移动大数据平台践行之路(上) ■ 浙江移动大数据平台践行之路(下) ■ 重读《大数据时代》:关于大数据的再认识 ■ 天龙八步:传统企业大数据运营的一些思考 ■ 七剑下天山, 谈谈我认识的精准营销 ■ 涅槃?高效报表开发人员的五件武器



傅一平 博士 毕业于浙江大学 从事电信行业工作,专注于大数据采集、处理、建模、管理、变现及

如果小伙伴需要在其它公众号转载这篇文章,在转载之前请通过以下邮箱告知。我欢迎大家转载,

作者简介

产业等研究

版权申明

但希望劳动成果获得大家的尊重。 邮箱: fuyp@zj.chinamobile.com