

企业数据科学修炼指南

—— 2018年版



云计算的实现，让我们迎来了大数据时代。与此同时，与大数据相关的职业也日益火爆。我们意识到，数据分析能力已成为企业的核心竞争力之一。然而，“数据分析”究竟是怎样一种能力？美国大数据专家Bernard Marr认为，有六项能力，是这其中最基本的、不可或缺的技能模块。

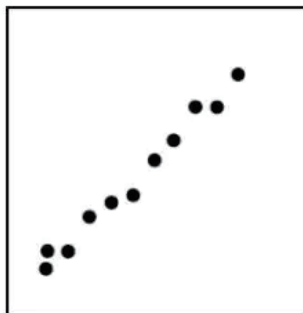
分析能力

在六项基本能力之中，首当其冲的就是分析能力——这项能力，可以帮你理解你所拥有的那些海量数据。

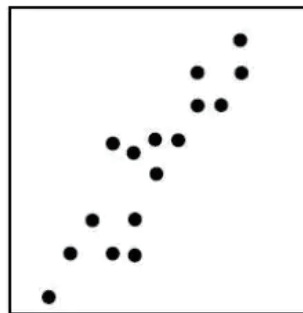
它包括：

- 确定与问题具有相关性的数据的能力；
- 解读数据、获得答案的能力。

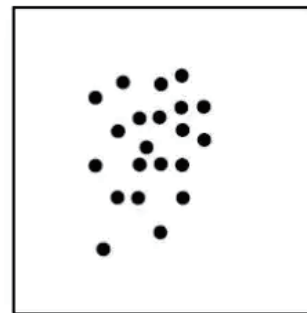
如果你善于发现模式，在因果之间建立联系，在帮助你把“企业数据”转化为“可操作的计划”方面，这些能力将非常宝贵。



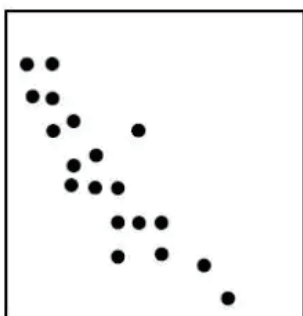
Strong positive correlation



Moderate positive correlation



No correlation



Moderate negative correlation



Strong negative correlation



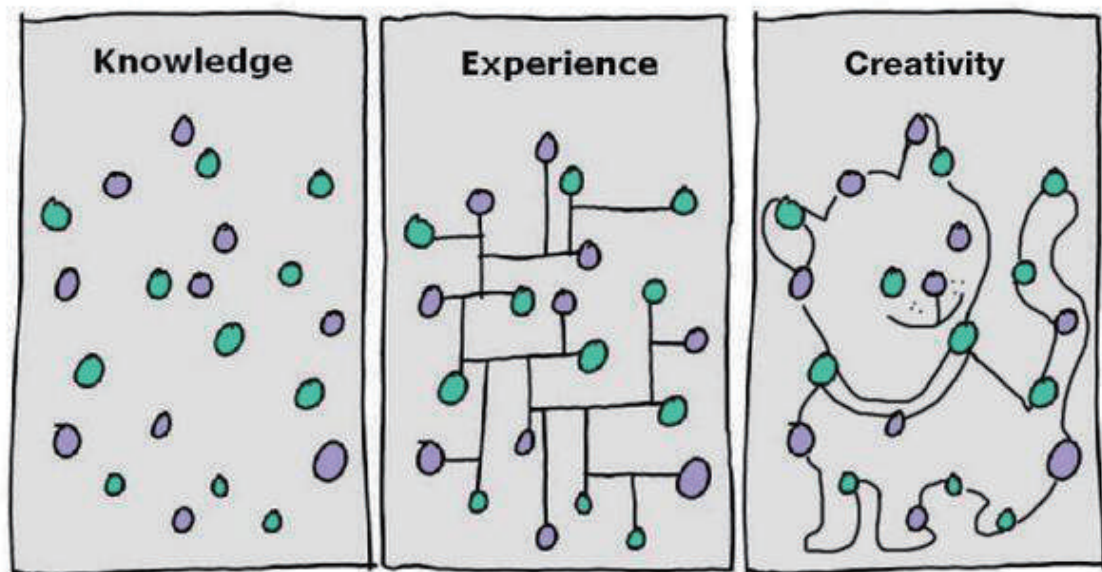
Curvilinear relationship

在“发现模式”方面，可视化是非常有力的工具

创造力

数据科学是一个新兴领域，对于企业应当如何使用数据没有硬性规定，也没有捷径可循。这意味着，具备“不断提出新的数据收集、解释、分析以及提取策略方法的能力”，将十分有价值。

未来，企业的“数据明星”将是那些有能力找到新数据来解决业务问题、并提出新的数据应用分析方法的人。

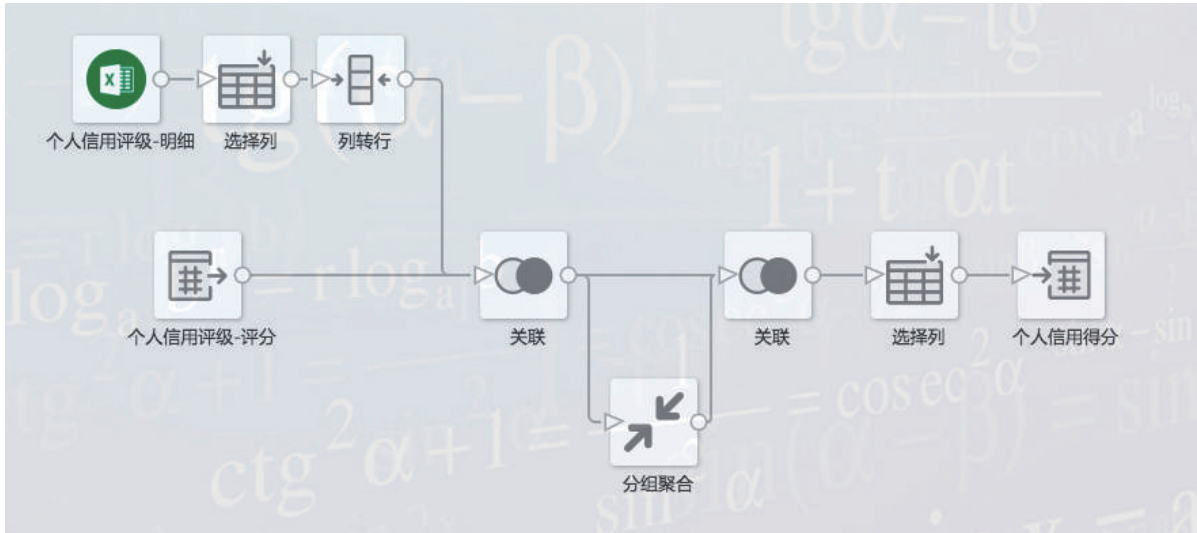


《知识、经验与创造力》经验可以统合知识，创意则可以让知识焕发新生命
作者：Sebastian Yuen

数学和统计学

优质、“老派”的数据处理方式依旧重要——也就是作为基础的数学与统计学。尽管在大数据战略中，非结构化数据的比例越来越大，但是，大量被收集、存储、并准备进行分析的信息，依然采用的是数字的形式。

即使在处理非结构化数据时，目标也往往是将数据元素（如邮件、社交媒体信息等等）简化至可以量化的数字，以便从中得出明确的结论。这意味着具备强大的数学或统计学背景，将有助于企业更好地拥抱大数据。



使用ETL工具对数据进行采集、转换与存储，背后也离不开数学与统计学的支持

计算机科学

计算机是大数据策略背后的重要工具，而程序员则需要拿出算法，把“数据”处理成“洞察”。这其中涵盖了机器学习、数据库或云计算等各个子领域，这些知识将成为新兴数据科学家们的有力武器。

开源技术尤其应当引起注意——比如Hadoop、Python、Apache Pig等——它们构成了大多数数据科学项目的基础。



业务技能

了解业务目标是什么、如何用流程驱动业务增长，对于数据分析师来说也十分重要。

一个十分危险的想法——公司应该雇佣一位聪明绝顶、理论知识丰富的数据科学家，把他供养在布满了服务器的地下数据实验室，而他将通过一个暗门向全公司传送自己的发现。

但是，任何一个与数据打交道的人，都不应当“闭门造数”，而是应与公司业务融为一体，牢牢掌握公司的业务目标，理解什么是朝着正确的方向前进、在前进的过程中，哪些数据才能彼此构成为“关键绩效指标”——那一定是基于对行业、企业、业务这“三业”的深刻理解而制定的。

数据分析始终服务于业务增长，也因此，现在有越来越多的技术 / 产品，都是为了实现“让业务人员具备数据分析能力”而诞生。



“让业务人员具备数据分析能力”的新一代商业智能平台

沟通能力

沟通技巧对于数据科学家来说必不可少。他们的分析结果至少要传达给两种人：关键决策者、团队其它成员。而这两种人通常都不能直接理解数据。

这时，将数据可视化、用数据“讲故事”的技能就至关重要了。你可以拥有世界上最好的分析能力，但除非你能使你的研究成果被人理解，并展示这些结果将如何帮他们提高绩效、推动成功，否则这些结果对企业来说将一文不值。



数据视觉功能，将数据变成可以普通人也可以看懂的故事
 大大提升数据分析师的沟通能力

