

数据科学入门及 R 语言简介

李伟

目录

| | |
|-------------|---|
| 数据科学在商业中的作用 | 2 |
| 案例一 | 3 |
| 案例二 | 5 |
| R 语言 | 6 |
| 课程安排 | 7 |
| 学习必备 | 8 |
| 参考书目 | 9 |

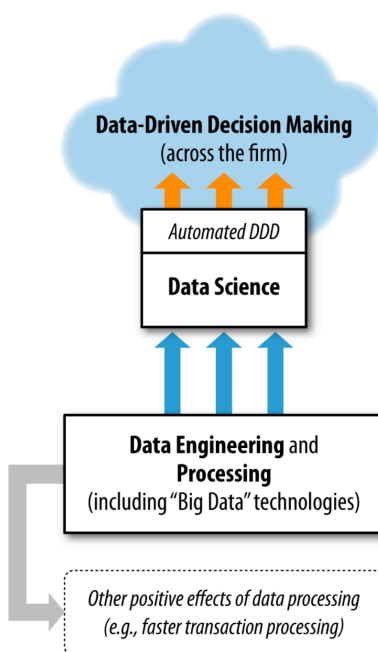


图 1: 数据科学

数据科学在商业中的作用

Dream no small dreams for they have no power to move the hearts of men.(Johann Wolfgang von Goethe)

数据科学在商业中的终极目标是辅助商业决策，提高决策的有效性。

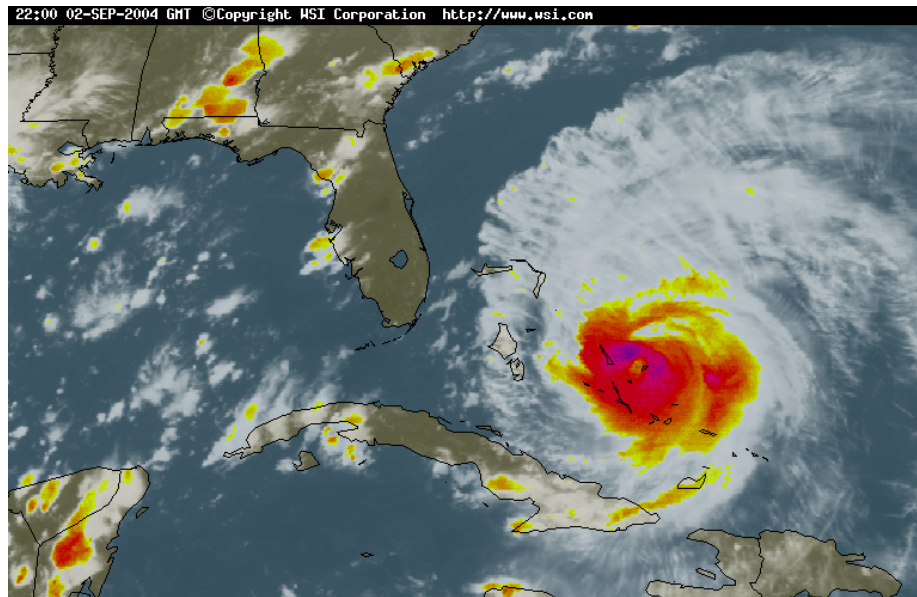


图 2: 飓风 Frances

案例一

2004 年，飓风 Frances 一周后即将袭击美国 Florida，此时，Wal-Mart 的 CIO, Linda M. Dillman, 要求其员工分析一下几周前飓风 Charley 来临时的销售数据，利用其数据库中的海量数据，她希望未雨绸缪。

最基本的需求是分析不同物品的需求量，从而为飓风来临准备充足的产品。

通过分析，发现飓风来临前销量最高的是啤酒，同时 strawberry Pop-Tarts 的销量增长了七倍！

案例二

Dickey's Barbecue Pit 是一家拥有超过 500 家连锁店的餐厅, 主营烧烤。其 CEO 采用数据分析之后, 通过以下四个方面改进了餐厅经营情况。

1. 数据说明, 其中午主要的顾客是驾驶 SUV 去上班的中年男子, 与最近的连锁店的车程大约 30 分钟。从而他们针对每个店附近 15 分钟到 30 分钟的 Ford 车顾客进行营销;
2. 数据说明, 在每个周三的下午, 会有母亲带着小孩来到店里享受午餐, 在餐厅呆很久。从而他们开展了一个“周三手工活动”, 以此来吸引更多的顾客;
3. 数据说明, 顾客非常喜欢 Fantasy Football 与狗。从而他们在 Fantasy Football 与狗的网站上做了大量的广告。
4. 建立数据分析平台, 展示实时数据, 从而管理层可以更好地进行决策。比如根据库存的多少进行相应的打折促销等。



图 3: R 语言

R 语言

R 语言是做统计的不二之选。主要基于以下两个理由:

1. R 语言是免费开源的;
2. R 语言拥有众多的软件包 (2019 年 11 月 14 日共有 15243 个包), 几乎可以解决你所有的问题, 很多软件包的作者都是统计界的资深人士。

首先, 安装 R 语言。下载地址是 www.r-project.org;

其次, 安装 Rstudio. 下载地址是 www.rstudio.com/products/rstudio/download/

课程安排

目前初步计划讲授的内容以“R 神”Hadley Wickham 的 R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 这本书为主, 网址是 <https://r4ds.had.co.nz/>, 中文版称为“R 数据科学”。

这本书学过后, 就可以胜任信息部的基本工作了! 培训后会有考试。考试通过可以来信息部工作! 我们信息部热烈欢迎!

学习必备

1. 笔记本电脑, 内存 4G, 最好是 8G, 操作系统 Windows7 以上, 最好是 Window10; 并且在其上安装 R 软件与 Rstudio.
2. R for Data Science, 中文版或者英文版, 推荐英文版;

参考书目

Provost, Foster, and Tom Fawcett. 2013. *Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking*.

Wickham, Hadley, and Garrett Grolemund. 2016. *R for Data Science : Import, Tidy, Transform, Visualize, and Model Data*.