

17Z BOOKS  
华章科技

网站分析领域的开创性著作，首度将六西格玛管理工具和思想引入网站分析全流程，使分析思维和方法更科学化和系统化

全面讲解网站分析必备的各项知识和技术，充分讲解Excel在网站分析中的各种功能和用法，包含大量案例和最佳实践



技术丛书



Website Analysis: Website Management, Optimization, and Operations Driven by Data

# 网站数据分析

## 数据驱动的网站管理、优化和运营

张洪举◎著



机械工业出版社  
China Machine Press



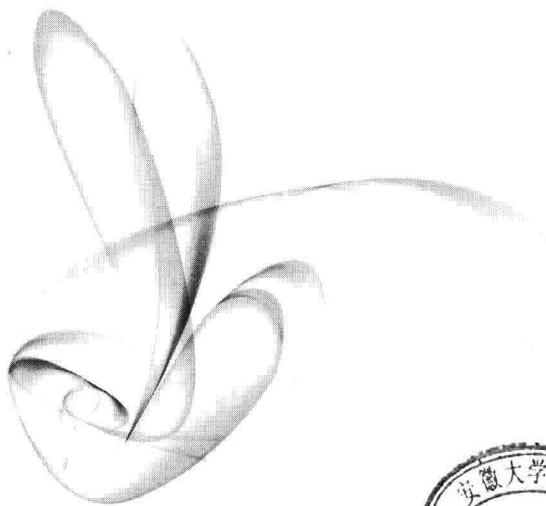
技术丛书

Website Analysis: Website Management, Optimization, and Operations Driven by Data

# 网站数据分析

## 数据驱动的网站管理、优化和运营

张洪举◎著



机械工业出版社  
China Machine Press

## 图书在版编目 (CIP) 数据

---

网站数据分析: 数据驱动的网站管理、优化和运营 / 张洪举著. —北京: 机械工业出版社, 2013.8  
(大数据技术丛书)

ISBN 978-7-111-43514-3

I. 网… II. 张… III. 网站—数据管理 IV. TP393.092

中国版本图书馆 CIP 数据核字 (2013) 第 172330 号

---

### 版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问 北京市展达律师事务所

网站分析领域的开创性著作, 首度将六西格玛质量管理工具和思想引入网站分析中, 将网站分析的复杂过程统一了起来, 使分析的思维和方法更科学化和系统化, 这是质量管理思想和数据分析技术的伟大结合。与此同时, 本书将网站分析必备的各项方法和技术, 充分融入到了整个体系中, 包含大量案例和最佳实践, 真正做到了用数据来驱动网站的管理、优化和运营。

全书一共分为五个部分: 第一部分 (第 1~2 章), 简要介绍了网站分析和六西格玛质量管理的主要内容, 以及 Excel 软件的主要功能和进行数据分析所应掌握的基本知识; 第二部分 (第 3 章) 是六西格玛改进模型的定义阶段, 讲解了如何识别和描述问题、理清问题边界、收集客户需求和确定客户需求的关键质量特性等; 第三部分 (第 4~8 章) 是六西格玛改进模型的测量阶段, 讲解了流程测量的起点、网站常用度量指标, 以及 Excel 基本数据处理技巧和数据透视表、SQL 查询等高级处理技巧; 第四部分 (第 9~11 章) 是六西格玛改进模型的分析阶段, 介绍了流程分析、逻辑树分析、杜邦分析、劣质成本分析、统计图表分析等多种分析方法, 以及如何在数据分析的基础上查找问题的根本原因; 第五部分 (第 12~13 章) 是六西格玛改进模型的改进和控制阶段, 介绍了如何确定改进方案和确保各项改进措施能够得到有效执行。

机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码 100037)

责任编辑: 陈佳媛

蕻城市京瑞印刷有限公司印刷

2013 年 8 月第 1 版第 1 次印刷

186mm×240mm·24.25 印张

标准书号: ISBN 978-7-111-43514-3

定 价: 69.00 元

---

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzsj@hzbook.com

## 为什么写作本书

自 2012 年 4 月开始动笔到最终定稿，本书的编写花费了我近一年的时间。写作是一件非常熬人的事情，繁忙的工作之余，我通常只能在晚上奋笔疾书，好在是兴趣所在，倒也不觉得辛苦，怡然自得其乐。

这是一本融合了 Excel 数据分析、六西格玛质量管理和网站分析艺术的图书。

毋庸置疑，Excel 是办公一族的首选数据处理和分析软件，市面上有许多介绍 Excel 操作的图书，但是能够与实际的企业管理和业务分析结合起来讲解的并不多见。六西格玛质量管理是一种成熟的企业管理和流程优化工具，被国内外许多大型公司采用，但是将其应用于电子商务网站管理的可谓凤毛麟角。网站分析是一个全新的行业，Google Analytics 等网站数据挖掘工具如雨后春笋般快速发展，但是它们关注的重点通常是网页流量和网页内容，能够联系企业整体情况进行分析的可谓少之又少。本书的写作目的就是三者进行有效的融合，在进行网站分析时，能够从“以客户为中心”的角度出发，借助 Excel 和六西格玛工具，全方位地分析问题原因、提出优化改进建议，而不仅仅是局限于网页表面文章。

## 读者对象

本书的内容较为广泛，但是并不复杂，任何对数据分析、网站分析、六西格玛质量管理感兴趣的朋友都可以阅读，可以通过具体的应用来提升企业管理水平和自己的数据分析能力。

本书的主要读者对象包括：

- 网络营销与运营管理人员
- 企业市场营销分析人员
- 企业财务管理人员
- 企业质量管理人员

- 企业部门经理与企业负责人
- 从事咨询、研究、分析的专业人士
- 计算机专业或市场营销、企业管理专业的学生

## 如何阅读本书

本书是按照六西格玛改进模型的定义、测量、分析、改正、控制步骤进行讲述的。如果读者初次接触六西格玛质量管理，建议从头开始阅读，深入了解解决问题的思路和方法；如果读者仅仅是希望掌握一些 Excel 数据分析技巧，可以选择感兴趣的章节进行阅读；如果读者希望学习 Google Analytics，很抱歉这本书无法满足您的要求，网站流量数据仅仅是网站分析数据的一个方面，尤其是对于电子商务网站，许多分析数据可能直接来自于企业内部数据库，而无需从外部获取。

本书内容大致可以分为以下五个部分。

第一部分是基础知识介绍，即第 1 章和第 2 章，简要介绍了网站分析和六西格玛质量管理的主要内容，以及 Excel 软件的主要功能和进行数据分析所应掌握的基本知识。

第二部分是六西格玛改进模型的定义阶段，即第 3 章，介绍了如何识别和描述问题、理清问题边界、收集客户需求和确定需求的关键质量特性等。

第三部分是六西格玛改进模型的测量阶段，即第 4 章～第 8 章，介绍了流程测量的起点、网站常用度量指标，以及 Excel 基本数据处理技巧和数据透视表、SQL 查询等高级处理技巧。

第四部分是六西格玛改进模型的分析阶段，即第 9 章～第 11 章，介绍了流程分析、逻辑树分析、杜邦分析、劣质成本分析、统计图表分析等多种分析方法，以及如何在数据分析的基础上查找产生问题的根本原因。

第五部分是六西格玛改进模型的改进和控制阶段，即第 12 章和第 13 章，介绍了如何确定改进方案以及如何确保各项改进措施能够得到有效执行。

## 勘误和支持

由于本人水平有限，书中错误在所难免，恳请您在阅读时给予批评和指正。您可以访问我的博客 (<http://blog.csdn.net/zhanghongju>) 和我交流，也可以直接将您的宝贵意见发送到 [sddzzhj@163.com](mailto:sddzzhj@163.com)。我将及时修改书中的错误并在我的博客中发布，同时也会尽最大努力为您提供支持和帮助，期待您的反馈。

本书中使用的 Excel 版本为 Excel 2010，Minitab 版本为 Minitab 16。

## 致谢

在本书的编写过程中，许多朋友给予了无私的支持与帮助，并参与了本书部分章节的资料搜集和整理工作，他们是凌亚东、魏兰花、谢景文、王欣、姜建友、姜岚、王勇、张雪梅等，是众人的经验与奉献成就了此书！

感谢机械工业出版社华章公司的策划编辑杨福川老师，与他的合作总是十分愉快！感谢姜影老师孜孜不倦地帮我修改稿件！

感谢我的家人，谢谢你们的支持与陪伴！

张洪举

# 目 录

---

前言

## 第 1 章 何谓网站分析.....1

1.1 网站分析不神秘.....1	1
1.1.1 网站分析的必要性.....2	2
1.1.2 网站分析的内容.....2	2
1.1.3 网站分析秘笈.....3	3
1.2 网站分析的流程.....4	4
1.2.1 什么是六西格玛质量管理.....5	5
1.2.2 六西格玛改进模型.....6	6
1.2.3 六西格玛在网站分析中的应用.....7	7
1.3 使用 Excel 进行数据分析.....8	8
1.3.1 数据记录.....8	8
1.3.2 数据计算.....8	8
1.3.3 数据分析.....9	9
1.3.4 数据图表.....9	9
1.3.5 协同工作.....10	10
1.3.6 编程开发.....11	11
1.4 本章小结.....11	11

## 第 2 章 分析师必备的数据知识.....12

2.1 数据类型.....12	12
2.1.1 数值类型.....12	12
2.1.2 文本类型.....13	13
2.1.3 日期和时间类型.....13	13
2.1.4 公式类型.....15	15
2.1.5 错误值类型.....15	15

2.1.6	逻辑值类型	17
2.1.7	数组类型	17
2.2	数字格式	17
2.2.1	可用的数字格式	17
2.2.2	自定义数字格式设置	19
2.3	数据类型与数字格式	22
2.3.1	如何定义数据类型	22
2.3.2	如何判断数据类型	22
2.3.3	数据类型的转换	23
2.4	公式与函数	27
2.4.1	运算符与优先级	27
2.4.2	单元格的引用方式	28
2.4.3	函数	31
2.4.4	数组公式	33
2.4.5	定义名称	33
2.5	本章小结	36

## 第3章 如何查找网站存在的问题 37

3.1	获取问题的途径	37
3.1.1	外部客户之声	37
3.1.2	内部用户之声	37
3.2	确定工作任务	39
3.2.1	二八定律	39
3.2.2	确定需解决的问题	40
3.2.3	确定改进目标	42
3.2.4	确定改进范围	42
3.3	收集客户需求	44
3.3.1	识别目标客户	44
3.3.2	客户需求调查	45
3.3.3	客户需求整理	49
3.4	确定关键质量特性	51
3.4.1	设置规格限	51
3.4.2	确定主要关键质量特性	52
3.5	利益相关者分析	53
3.5.1	分析的内容及作用	53



3.5.2 确定关键利益相关者	53
3.6 本章小结	54

## 第4章 如何测量问题现状 56

4.1 如何详细描述流程	56
4.1.1 流程图的作用	56
4.1.2 流程图的分类	57
4.1.3 绘制流程图	57
4.2 如何进行流程测量	60
4.2.1 确定测量指标	61
4.2.2 确定测量对象	61
4.3 流程现状与差距分析	62
4.3.1 平均值与标准差	62
4.3.2 西格玛水平	63
4.3.3 流通合格率	64
4.3.4 差距分析	65
4.4 网站常用度量指标	65
4.4.1 唯一 IP	66
4.4.2 唯一访问者	66
4.4.3 页面浏览量与唯一页面浏览量	67
4.4.4 页面驻留时间与网站驻留时间	67
4.4.5 跳出率	68
4.4.6 广告显示数、点击数和点击率	68
4.4.7 转换率	68
4.4.8 回访者比率	68
4.4.9 积极访问者比率	69
4.4.10 忠实访问者比率	69
4.5 本章小结	69

## 第5章 数据输入技巧 70

5.1 数据录入前奏	70
5.1.1 格式化窗口显示	70
5.1.2 格式化单元格显示	76

5.1.3	设置数字格式与数据类型	80
5.2	基本数据输入技巧	80
5.2.1	自动输入小数点	80
5.2.2	输入分数	81
5.2.3	输入日期和时间	82
5.2.4	输入指数	83
5.2.5	强制换行操作	84
5.2.6	改变回车移动方向	84
5.2.7	使用记录单输入数据	85
5.3	规律数据输入技巧	86
5.3.1	多单元格同时输入相同数据	86
5.3.2	序列自动填充	87
5.3.3	记忆式输入	89
5.3.4	使用列表选择	91
5.4	数据有效性	91
5.4.1	数据有效性的设置内容	91
5.4.2	设置数值允许范围	93
5.4.3	设置日期和时间允许范围	93
5.4.4	设置文本长度允许范围	94
5.4.5	设置允许的序列值	95
5.4.6	自定义有效性设置	95
5.4.7	定位、复制和删除数据有效性	96
5.5	复制、剪切与粘贴	98
5.5.1	复制与剪切操作	98
5.5.2	粘贴操作	99
5.5.3	使用 Office 剪贴板简化操作	103
5.6	本章小结	104
<b>第 6 章</b>	<b>基本数据处理</b>	<b>105</b>
6.1	数据修正	105
6.1.1	文本联接	105
6.1.2	文本分列	106
6.1.3	文本截取	109
6.1.4	文本规范	111
6.1.5	数字格式化	112

6.1.6 精确比较文本	113
6.1.7 字符代码转换	113
6.1.8 数据的行列转换	114
6.2 数据标准化	115
6.2.1 数据缩放处理	115
6.2.2 指标方向调整	116
6.2.3 设置指标权重	116
6.3 数据检索与组合	116
6.3.1 使用 VLOOKUP 检索	117
6.3.2 使用 HLOOKUP 检索	117
6.3.3 使用 MATCH 检索	118
6.4 数据分组	119
6.5 数据排序	119
6.5.1 按列排序	120
6.5.2 按行排序及排序选项	121
6.5.3 按自定义序列排序	122
6.6 重复数据处理	123
6.6.1 使用函数识别重复数据	124
6.6.2 使用条件格式标识重复数据	124
6.6.3 使用筛选剔除重复数据	125
6.6.4 直接删除重复数据	126
6.7 数据筛选	126
6.7.1 自动筛选	127
6.7.2 高级筛选	129
6.8 数据抽样	134
6.8.1 使用“抽样”工具抽样	134
6.8.2 使用 RAND 函数抽样	135
6.9 数据计算	136
6.9.1 计数	136
6.9.2 求和	139
6.9.3 计算平均值	141
6.9.4 计算频率	143
6.9.5 计算极值	145
6.9.6 计算排名	147
6.10 数据分类汇总	149
6.10.1 SUBTOTAL 函数	149

6.10.2 分类汇总工具	150
6.11 本章小结	153
<b>第 7 章 使用数据透视表进行数据处理</b>	<b>154</b>
7.1 创建数据透视表	154
7.1.1 创建基本数据透视表	154
7.1.2 创建动态数据透视表	156
7.1.3 以其他工作簿数据创建数据透视表	157
7.1.4 以文本文件数据创建数据透视表	159
7.2 布局数据透视表	161
7.2.1 推迟布局更新	161
7.2.2 改变字段名称	162
7.2.3 汇总字段显示设置	163
7.3 刷新数据	164
7.3.1 使用 Excel 数据源时自动刷新	164
7.3.2 引用外部数据源时自动刷新	165
7.3.3 链接数据透视表的刷新	165
7.4 数据筛选	166
7.4.1 了解筛选	166
7.4.2 手动筛选	166
7.4.3 使用行标签筛选	168
7.4.4 使用值筛选	169
7.4.5 使用切片器筛选	171
7.5 数据排序	173
7.5.1 手动排序	173
7.5.2 自动排序	174
7.6 项目组合	175
7.6.1 组合指定项	175
7.6.2 组合数字项	176
7.6.3 组合时间项	178
7.6.4 取消组合	180
7.7 数据计算	181
7.7.1 改变字段的汇总方式	181
7.7.2 插入计算字段和计算项	182
7.8 创建基于多表的数据透视表	185

7.8.1 添加“数据透视表和数据透视图向导”按钮到快速访问工具栏	185
7.8.2 创建基于单页字段的数据透视表	186
7.8.3 创建基于自定义页字段的数据透视表	187
7.9 本章小结	188

## 第 8 章 使用 SQL 查询进行数据处理 189

8.1 获取外部数据	189
8.1.1 获取 Excel 工作表数据	189
8.1.2 获取 Access 表格数据	191
8.1.3 获取网站表格数据	193
8.1.4 使用 OLEDB 方式连接数据库	194
8.1.5 使用 ODBC 方式连接数据库	196
8.1.6 获取 XML 文件中的数据	199
8.1.7 获取文本文件中的数据	200
8.2 基本的 SELECT 语句	202
8.2.1 SELECT 语句的结构	202
8.2.2 表的引用规则	204
8.3 使用选择列表和表别名	205
8.3.1 选择所有列	205
8.3.2 选择特定列	205
8.3.3 在选择列表中使用常量、函数和表达式	206
8.3.4 使用表别名	208
8.4 使用 WHERE 子句筛选行	208
8.4.1 使用比较搜索条件	208
8.4.2 使用范围搜索条件	209
8.4.3 使用列表搜索条件	209
8.4.4 使用模式匹配搜索条件	209
8.4.5 使用 NULL 比较搜索条件	211
8.5 使用 GROUP BY 子句和聚合函数进行分组计算	211
8.6 使用 HAVING 子句从分组后结果中筛选行	212
8.7 使用 ORDER BY 子句进行排序	213
8.7.1 指定排序列	213
8.7.2 指定排序顺序	215
8.8 使用 TOP 子句限制结果集	215

8.9 使用 DISTINCT 消除重复行	215
8.10 使用子查询	216
8.10.1 在选择列表中使用子查询	216
8.10.2 含有 IN 和 EXISTS 的子查询	217
8.10.3 含有比较运算符的子查询	220
8.10.4 使用 ANY、SOME 或 ALL 关键字	221
8.10.5 使用多层嵌套子查询	222
8.11 使用联接查询	223
8.11.1 联接的语法格式	223
8.11.2 列名限定和选择列表的使用	224
8.11.3 联接条件设定	224
8.12 使用交叉联接	225
8.12.1 交叉联接的语法格式	225
8.12.2 使用交叉联接提高查询效率	226
8.12.3 为交叉联接添加 WHERE 子句	227
8.13 使用内部联接	227
8.13.1 内部联接的语法格式	228
8.13.2 等值内部联接	228
8.13.3 不等值联接	229
8.14 外部联接	230
8.14.1 外部联接的语法格式	230
8.14.2 左外部联接	231
8.14.3 右外部联接	232
8.14.4 完全外部联接	233
8.15 使用自联接	234
8.15.1 使用不同列实现自联接	234
8.15.2 使用同一列实现自联接	235
8.16 合并结果集	235
8.16.1 使用 UNION 与 UNION ALL 进行结果集合并	236
8.16.2 使用 ORDER BY 子句对合并结果集排序	236
8.16.3 结果集的合并顺序	237
8.17 本章小结	237
<b>第 9 章 过程与数据分析</b>	<b>239</b>
9.1 过程分析	239

9.1.1	绘制详细流程图	239
9.1.2	逻辑树分析法	241
9.1.3	杜邦分析法	242
9.2	减少流程中的浪费	243
9.2.1	精益六西格玛与传统六西格玛的区别	244
9.2.1	区分增值和非增值活动	244
9.3	劣质成本分析	248
9.3.1	劣质成本分析的作用	248
9.3.2	劣质成本的构成	249
9.3.3	劣质成本分析流程	251
9.3.4	劣质成本分析的注意事项	252
9.4	流通合格率分析	253
9.4.1	流通合格率分析的作用	253
9.4.2	流通合格率的计算方法	254
9.5	统计图表分析	254
9.5.1	正确选择统计图表	254
9.5.2	统计图表的结构与调整	258
9.5.3	直方图(频率分布)分析	265
9.5.4	箱线图(数据分布)分析	268
9.5.5	时间序列图(趋势)分析	273
9.5.6	散点图(相关性及数据分布)分析	274
9.5.7	对比图分析(差异分析)	279
9.5.8	算术平均分析(差异分析)	284
9.5.9	移动平均分析(趋势分析)	287
9.5.10	漏斗图分析(差异分析)	290
9.6	本章小结	294

## 第 10 章 查找问题根本原因 295

10.1	广泛寻找问题原因	295
10.1.1	鱼骨图分析	295
10.1.2	5W2H 分析	298
10.2	识别问题根本原因	299
10.2.1	帕累托图分析	299
10.2.2	因果矩阵分析	304

10.2.3	5个为什么分析	306
10.3	本章小结	307
<b>第 11 章</b>	<b>分析结果展现</b>	<b>308</b>
11.1	表格展现策略	308
11.1.1	使用突出显示	308
11.1.2	使用数据条	309
11.1.3	使用色阶	310
11.1.4	使用图标集	310
11.1.5	使用迷你图	311
11.2	图表展现策略	313
11.2.1	信息全面	313
11.2.2	适度简约	316
11.2.3	字体统一	319
11.2.4	颜色协调	320
11.2.5	重点突出	321
11.2.6	顾全大局	327
11.3	数据分析报告	328
11.3.1	撰写要求	328
11.3.2	结构与内容	328
11.4	本章小结	332
<b>第 12 章</b>	<b>确定解决方案</b>	<b>333</b>
12.1	提出解决方案	333
12.2	评估解决方案	334
12.2.1	资格标准评估	334
12.2.2	参考标准评估	334
12.3	执行解决方案	337
12.3.1	失效模式及后果分析	337
12.3.2	制定沟通计划	338
12.3.3	培训及试运行	339
12.4	本章小结	339



<b>第 13 章 监控改进后的内容</b> .....	340
13.1 建立控制计划.....	340
13.2 使用控制图监控改进效果.....	341
13.2.1 控制图的样式.....	341
13.2.2 控制图的类型.....	342
13.2.3 常规异常判断原则.....	343
13.2.4 p 图——监测客户转化率.....	346
13.2.5 X-MR 图——监测客单投递时长.....	350
13.3 流程标准化建设.....	353
13.3.1 标准化建设的作用.....	354
13.3.2 良好标准的制定要求.....	354
13.3.3 标准的制定步骤.....	354
13.3.4 标准作业程序的格式.....	355
13.4 本章小结.....	356
<b>附录 A Google Analytics 简介</b> .....	357
<b>附录 B Minitab 简介</b> .....	361
<b>附录 C Microsoft Visio 简介</b> .....	364
<b>附录 D 网站分析案例——客户价值分析</b> .....	365