

国内著名资深实战型Excel高级应用培训专家
倾力力作!

Power Query

智能化数据汇总与分析

- 让你能够高效处理数据的Excel技能技巧大全
- 使你快速成为数据分析高手的Excel精粹合集
- 帮你处理实际工作疑难杂症的Excel实战案例

为个人提升办公效率!
为企业提供解决方案!
打造Excel职场大百科!

韩小良◎著



中国水利水电出版社
www.waterpub.com.cn



Excel大百科全书

Power Query

智能化数据汇总与分析

韩小良◎著



中国水利水电出版社

www.waterpub.com.cn

·北京·

内 容 提 要

作为一本Power Query初级入门读本,《Power Query 智能化数据汇总与分析》不深入介绍M函数,而是结合大量的实际案例,利用Power Query可视化的操作向导,来解决实际工作中烦琐的数据整理、汇总和分析问题,让读者在短时间内,能够对Power Query有一个较为全面的了解和掌握,并能迅速提升数据处理和统计分析效率。

《Power Query 智能化数据汇总与分析》适合具有Excel基础知识并经常处理大量数据的各类人员阅读,也可作为大专院校经济类本科生、研究生和MBA学员的教材或参考书。

图书在版编目(CIP)数据

Power Query 智能化数据汇总与分析 / 韩小良著. —北京:
中国水利水电出版社, 2019.10

ISBN 978-7-5170-7788-6

I. ① P… II. ①韩… III. ①表处理软件 IV. ① TP391.13

中国版本图书馆CIP数据核字(2019)第131393号

书 名	Power Query智能化数据汇总与分析
作 者	Power Query ZHINENGHUA SHUJU HUIZONG YU FENXI 韩小良 著
出版发行	中国水利水电出版社 (北京市海淀区玉渊潭南路1号D座100038) 网址: www.waterpub.com.cn E-mail: zhiboshangshu@163.com 电话: (010) 62572966-2205/2266/2201 (营销中心)
经 售	北京科水图书销售中心(零售) 电话: (010) 88383994、63202643、68545874 全国各地新华书店和相关出版物销售网点
排 版	北京智博尚书文化传媒有限公司
印 刷	河北华商印刷有限公司
规 格	180mm×210mm 24开本 18.5印张 628千字 1插图
版 次	2019年10月第1版 2019年10月第1次印刷
印 数	0001—5000册
定 价	79.80元

凡购买我社图书,如有缺页、倒页、脱页的,本社营销中心负责调换
版权所有·侵权必究



前言

Preface

在每次培训课程上,总会有学生问:韩老师,如何把大量的工作表数据汇总到一张工作表上?遇到这样的提问,我会问以下几个问题。

(1)工作表规范吗?也就是说,每张工作表是否是标准规范的表单。第一行就是标题,没有合并单元格,没有大标题小注脚,没有垃圾数据。

(2)是当前工作簿里的几张工作表,还是多个工作簿里的多张工作表?

(3)你要汇总成什么样的表格?是把这些工作表数据简单地堆积到一张工作表中,还是要做一张统计分析表?

(4)表格数据量大吗?是不是数据会随时变化?

学生的回答是表格很规范,结构也一样,就是每个月都要做大量的复制、粘贴工作,累人,还容易出错。

Excel 2016的面世,将Excel的数据处理与数据分析提升到了一个新高度。不论是一个工作簿的多张工作表,还是多个工作簿的多张工作表;不论是打开的工作簿,还是没有打开的工作簿;不论是数据库数据,还是文本数据,或者是Excel工作簿数据,诸如此类的大量数据汇总与分析,在Excel 2016的新工具Power Query面前,已经不再是一件令人焦虑的事情了。你需要做的仅仅是把基础表单做规范,然后动动鼠标,用几个简单的命令,按照可视化的向导一步一步操作,即可快速完成。

对于大多数人来说,Power Query是一个陌生的工具,觉得很难学、很难掌握。其实,对于人们日常的数据处理和统计分析来说,掌握Power Query的主要使用方法和实际应用就足够了,毕竟,绝大部分人不是专业的数据分析师,也不需要建立多么复杂的数据模型来开发高端的商业智能。

因此,本书不深入介绍对大多数初学者来说难懂的M语言,不介绍在实际工作中用途不大的表理论,而是应用Power Query来解决实际工作中烦琐的数据整理、汇总和分析问题。只为解决问题,帮读者快速提升数据处理效率,从烦人、累人的数据处理工作中解放出来,



是本书的宗旨。

本书共分9章,结合大量的实际案例,介绍Power Query在数据查询和汇总中的各种实际应用,提供详细的操作步骤,读者仔细阅读并实操,就能很快掌握Power Query实用技能。

本书的所有案例都在Excel 2016以上的版本中测试完成。

本书的编写得到了朋友和家人的支持与帮助,参与编写的人员有杨传强、于峰、李盛龙、董国灵、毕从牛、高美玲、王红、李满太、程显峰、王荣亮、韩良智、韩舒婷、翟永俭、贾春雷、冯岩、韩良玉、徐沙比、申果花、韩永坤、冀叶彬、刘兵辰、徐晓斌、刘宁、韩雪珍、徐换坤、张合兵、徐克令、张若曦、徐强子等,在此表示衷心的感谢!

中国水利水电出版社的刘利民老师和秦甲老师也给予了很多帮助与支持,使得本书能够顺利出版,在此表示衷心的感谢。

由于认知有限,作者虽尽职尽责,以期本书能够满足更多人的需求,但难免有疏漏之处,敬请读者批评、指正。欢迎加入QQ群一起交流,QQ群号为580115086。

韩小良

Contents





01

Chapter

初识Power Query /1

1.1 两个实际案例引发的方法变革 /2

1.1.1 案例一：16个分公司全年工资汇总 /2

1.1.2 案例二：几年来近百万行销售数据统计分析 /3

1.2 Power Query来救命了 /4

1.2.1 可以从不同的数据源采集数据 /5

1.2.2 大量工作表数据汇总变得易如反掌 /5

1.2.3 以数据模型为核心的海量数据分析很简单 /5

1.2.4 体验前所未有的智能化操作 /5

1.3 Power Query命令与编辑器界面 /6

1.3.1 Power Query命令 /6

1.3.2 Power Query编辑器 /7

02

Chapter

Power Query的基本用法与 注意事项 /9

2.1 从工作簿里查询数据 /10

2.1.1 从当前工作簿里查询数据 /10

2.1.2 从其他没有打开的工作簿里查询数据 /20

2.2 从文本文件里查询数据 /20

2.2.1 CSV格式文本文件 /20

2.2.2 其他格式文本文件 /25

2.3 从数据库查询数据 /28

2.3.1 从Access数据库查询数据 /29

2.3.2 从SQL Server数据库查询数据 /31

2.4 保存查询结果 /33

2.4.1 保存为表 /33

2.4.2 保存为数据透视表 /35

2.4.3 保存为数据透视图 /36

2.4.4 仅创建连接 /36

2.4.5 将数据添加到数据模型 /36

2.4.6 重新选择保存方式 /37

2.4.7 导出连接文件并在其他工作簿中使用现有查询 /37

2.5 注意事项 /41

2.5.1 自动记录下每个操作步骤 /41

2.5.2 注意提升标题 /44

2.5.3 注意设置数据类型 /45

2.5.4 编辑已有的查询 /47

2.5.5 刷新查询 /47

2.6 了解每个操作步骤及相应M公式 /49

2.6.1 源 /49

2.6.2 导航 /50

2.6.3 提升的标题 /51

2.6.4 更改的类型 /51

2.6.5 筛选的行 /52

2.6.6 排序的行 /53

2.6.7 其他操作 /54

03

Chapter

Power Query常规数据处理操作/55

3.1 打开“Power Query编辑器”窗口 /56

3.1.1 执行“查询”命令打开 /56

3.1.2 通过某个查询打开 /56

3.1.3 直接打开 /57

3.2 查询的基本操作 /58

3.2.1 预览查询 /58

3.2.2 重命名查询 /59

3.2.3 复制查询 /61

3.2.4 删除查询 /62

3.2.5 刷新查询 /63

3.2.6 设置查询说明信息 /63

3.2.7 显示/隐藏“查询&连接”窗格 /65

3.3 列的一般操作 /65

3.3.1 重命名列 /65

3.3.2 选择要保留的列 /65

3.3.3 删除不需要的列 /67



- 3.3.4 复制列 /67
- 3.3.5 移动列位置 /68
- 3.3.6 拆分列 /68
- 3.3.7 合并列 /81
- 3.3.8 透视列 /84
- 3.3.9 逆透视列 /97
- 3.3.10 替换列数据 /100
- 3.4 文本列的特殊操作 /102**
 - 3.4.1 从列数据中提取字符 /102
 - 3.4.2 转换列数据格式 /109
- 3.5 日期时间列的特殊操作 /113**
 - 3.5.1 计算当前与表格日期之间的天数 /114
 - 3.5.2 从日期和时间数据中提取日期 /115
 - 3.5.3 计算日期的年数据 /118
 - 3.5.4 计算日期的月数据 /119
 - 3.5.5 计算日期的季度数据 /121
 - 3.5.6 计算日期的周数据 /122
 - 3.5.7 计算日期的天数据 /123
 - 3.5.8 获取某列日期中的最早日期或最晚日期 /124
 - 3.5.9 合并日期和时间 /124
 - 3.5.10 处理时间列 /125
- 3.6 数字列的特殊操作 /125**
 - 3.6.1 对列数字进行批量修改 /126
 - 3.6.2 对列数字进行四舍五入 /128
 - 3.6.3 对列数字进行简单的统计计算 /128
 - 3.6.4 对列数字进行其他的计算处理 /128
- 3.7 数据行的一般操作 /129**
 - 3.7.1 保留行 /129
 - 3.7.2 删除行 /129
- 3.8 整个表的操作 /130**
 - 3.8.1 反转行 /130
 - 3.8.2 转置表 /131
 - 3.8.3 表标题设置 /132

04

Chapter

向表添加新列 /133

4.1 添加索引列 /134

4.1.1 添加自然序号的索引列 /134

4.1.2 添加自定义序号的索引列 /136

4.2 添加自定义列 /136

4.2.1 添加常数列 /137

4.2.2 添加常规计算列 /138

4.3 添加条件列 /142

4.3.1 添加条件列——结果是具体值 /142

4.3.2 添加条件列——结果是某列值 /147

4.3.3 删除某个条件 /150

4.3.4 改变各个条件的前后次序 /150

4.4 条件语句if then else /150

4.4.1 基本语法结构 /151

4.4.2 应用举例 /151

05

Chapter

查询分组统计 /159

5.1 基本分组 /160

5.1.1 对项目求和 /160

5.1.2 对项目求平均值、最大值和最小值 /163

5.1.3 对项目计数 /165

5.2 高级分组 /167

5.2.1 同时进行计数与求和 /167

5.2.2 同时进行计数、平均值、最大值和最小值 /169

5.2.3 对多个字段进行不同的分组 /172

5.2.4 删除某个分组 /175

5.2.5 调整各个分组的次序 /175

06

Chapter

多表合并查询 /176

6.1 一个工作簿内的多张工作表合并汇总 /177

- 6.1.1 多张工作表的堆积汇总 /177
- 6.1.2 多张工作表的关联汇总——两张工作表的情况 /183
- 6.1.3 多张工作表的关联汇总——多张工作表的情况 /191
- 6.1.4 多张工作表的关联汇总——匹配数据 /198

6.2 多个工作簿的合并汇总 /203

- 6.2.1 汇总多个工作簿,每个工作簿仅有一张工作表 /203
- 6.2.2 汇总多个工作簿,每个工作簿有多张工作表 /213
- 6.2.3 查找汇总多张工作表里满足条件的数据 /221
- 6.2.4 按项目分组汇总多张工作表的数据 /221

6.3 合并查询 /226

6.4 合并查询综合应用1 /234

- 6.4.1 只有一列需要核对的数据 /234
- 6.4.2 有多列需要核对的数据 /242

6.5 合并查询综合应用2 /246

- 6.5.1 建立基本查询 /247
- 6.5.2 统计全年在职员工 /248
- 6.5.3 统计全年离职员工 /249
- 6.5.4 统计全年新入职员工 /250

6.6 合并查询综合应用3 /251

- 6.6.1 建立基本查询 /251
- 6.6.2 统计两年存量客户 /252
- 6.6.3 统计去年的流失客户 /260
- 6.6.4 统计当年新增客户 /264

6.7 追加查询 /265

- 6.7.1 新建追加查询 /265
- 6.7.2 事后追加新的数据表 /270

6.8 其他合并问题 /274

- 6.8.1 核对总表和明细表 /274
- 6.8.2 制作已完成合同明细表 /279
- 6.8.3 制作未完成合同明细表 /286

07

Chapter

Power Query数据处理案例精粹 /288

7.1 拆分列 /289

- 7.1.1 “拆分列”命令 /289
- 7.1.2 按分隔符拆分列——拆分成数列 /289
- 7.1.3 按分隔符拆分列——拆分成数行 /296
- 7.1.4 按字符数拆分列 /298

7.2 合并列 /300

- 7.2.1 合并列形式1——合并为一列 /300
- 7.2.2 合并列形式2——合并为新列 /302

7.3 提取字符 /304

- 7.3.1 提取字符形式1——将原始列转换为提取的字符 /305
- 7.3.2 提取字符形式2——将提取的字符添加为新列 /306
- 7.3.3 提取最左边的字符 /306
- 7.3.4 提取最右边的字符 /308
- 7.3.5 提取中间字符 /309
- 7.3.6 提取分隔符之前的字符 /311
- 7.3.7 提取分隔符之后的字符 /313
- 7.3.8 提取分隔符之间的字符 /313
- 7.3.9 综合练习——从身份证号码中提取信息 /315

7.4 转换表结构 /320

- 7.4.1 一列变多列 /320
- 7.4.2 多列变一列 /320
- 7.4.3 一行变多行 /320
- 7.4.4 多行变一行 /321
- 7.4.5 二维表转换为一维表 /324
- 7.4.6 一维表转换为二维表 /326
- 7.4.7 综合练习1——连续发票号码的数据处理 /330
- 7.4.8 综合练习2——考勤数据处理 /342

7.5 表格合并 /348

- 7.5.1 汇总工作簿内工作表的两个重要问题 /348
- 7.5.2 一个工作簿的表格合并——全部工作表合并 /350
- 7.5.3 一个工作簿的表格合并——部分工作表合并 /355



7.5.4 不同工作簿的表格合并 /359

7.5.5 汇总多个文本文件 /369

7.6 表格查询 /373

7.6.1 单表查询满足条件的数据 /373

7.6.2 多表查询满足条件的数据 /376

7.7 基本统计汇总 /379

7.7.1 单列分组计算 /379

7.7.2 多列分组计算 /382

7.7.3 用透视列重构报表 /385

08

Chapter

与Power Pivot联合使用 /386

8.1 将Power Query查询加载为数据模型 /387

8.1.1 加载为数据模型的方法 /387

8.1.2 重新编辑现有的查询 /388

8.2 利用Power Pivot建立基于数据模型的数据透视表 /388

8.2.1 基于某一个查询的数据透视表 /388

8.2.2 基于多张有关联表查询的数据透视表 /393

8.2.3 基于海量数据查询的数据透视表 /397

09

Chapter

M语言简介 /401

9.1 从查询操作步骤看M语言 /402

9.1.1 查询表的结构 /402

9.1.2 每个操作步骤对应一个公式 /404

9.1.3 用高级编辑器查看完整代码 /404

9.2 通过手动创建行、列和表进一步了解M函数 /406

9.2.1 创建行 /406

9.2.2 创建列 /412

9.2.3 创建一个连续字母的列 /413

9.2.4 创建一个连续数字的列 /414

9.2.5 创建一个表 /414

9.3 M语言及函数 /415

9.3.1 M语言结构 /416

- 9.3.2 M语言的运算规则 /417
- 9.3.3 M函数语法结构 /417
- 9.3.4 M函数简介 /418
- 9.4 M函数应用举例 /420**
 - 9.4.1 分列文本和数字 /420
 - 9.4.2 从身份证号码中提取生日和性别 /424
 - 9.4.3 计算迟到分钟数和早退分钟数 /425

Chapter

01

初识Power Query

在介绍 Power Query 之前，先通过两个实际案例来看看对于诸如数据汇总问题，平时我们都是怎么做的，而使用 Power Query 又是怎么做的。

1.1

两个实际案例引发的方法变革

1.1.1 案例一：16 个分公司全年工资汇总

这是一个典型的实际问题：有 16 个工作簿，代表 16 个分公司；每个工作簿中有 12 张工作表，分别保存 12 个月的工资数据；每个分公司的工资表中，员工又分为合同工和劳务工。现在要求制作以下 3 张汇总表。

- (1) 全部分公司的合同工个税汇总表。
- (2) 全部分公司的劳务工个税汇总表。
- (3) 按分公司汇总社保、公积金和个税。

示例数据如图 1-1 和图 1-2 所示。

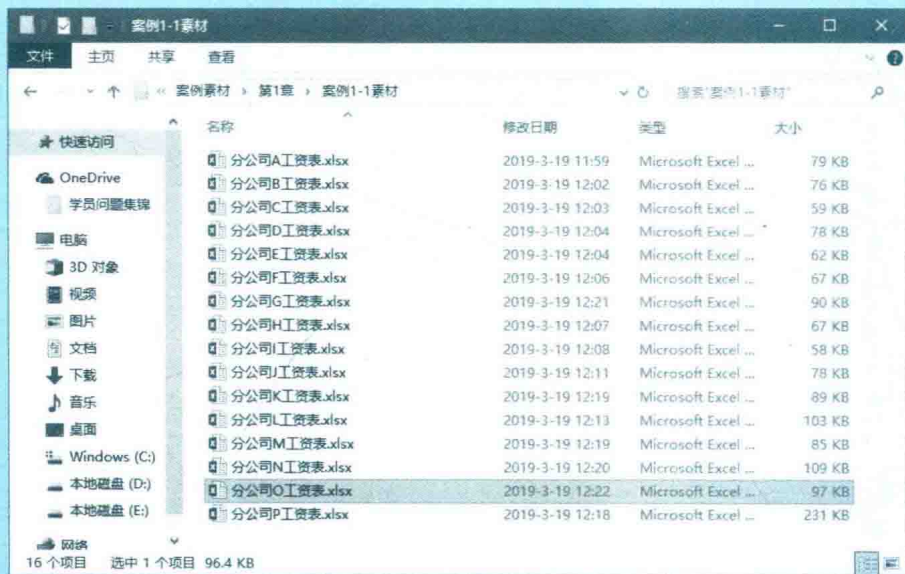


图1-1 文件夹里的16个工作簿

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	姓名	合同类型	基本工资	出勤工资	岗位津贴	福利津贴	应发工资	个人所得税	社保金	公积金	四金合计	实发工资	
2	C001	合同工	3027.00	162.00	192.00	139.00	3520.00	0.00	452.00	42.00	494.00	3026.00	
3	C002	合同工	7536.00	198.00	822.00	807.00	9363.00	331.30	294.00	296.00	590.00	8441.70	
4	C003	劳务工	6067.00	778.00	563.00	913.00	8321.00	227.10	209.00	65.00	274.00	7819.90	
5	C004	合同工	6912.00	315.00	742.00	589.00	8558.00	250.80	311.00	347.00	658.00	7649.20	
6	C005	合同工	7270.00	708.00	237.00	927.00	9142.00	309.20	409.00	200.00	609.00	8223.80	
7	C006	合同工	7326.00	60.00	848.00	536.00	8770.00	272.00	38.00	188.00	226.00	8272.00	
8	C007	合同工	4002.00	308.00	546.00	968.00	5824.00	24.72	11.00	63.00	74.00	5725.28	
9	C008	劳务工	10426.00	430.00	687.00	469.00	12012.00	847.40	199.00	346.00	545.00	10619.60	
10	C009	合同工	11334.00	49.00	351.00	356.00	12090.00	863.00	342.00	9.00	351.00	10876.00	
11	C010	合同工	8624.00	97.00	481.00	736.00	9938.00	432.60	277.00	63.00	340.00	9165.40	

图1-2 每个工作簿里的12张工作表

16个工作簿，每个工作簿里有12张工作表，总共 $16 \times 12 = 192$ 张工作表数据要汇总，这样的工作，大部分人就是将工作簿逐一打开，复制粘贴，然后制作数据透视表……3个小时就这样过去了。

掌握了Excel VBA技能的读者，马上想到了利用VBA来快速汇总，编写代码，调试代码，运行，检查，这样半个小时也过去了。

这也是没办法的事情，因为目前计算机上安装的是Office 2010这样的版本，只能这样做。

1.1.2 案例二：几年来近百万行销售数据统计分析

再来看这样一个例子。领导说，把2015年至今的销售数据做个分析，看看各项业务近年来的趋势及各个市场的变化，尤其是分析去年和今年的同比情况。

从ERP导出了2015年以来的销售数据，居然有近80万行，文件大小也高达数百兆。结果是此工作簿根本就无法正常工作，就连每保存一次，工作簿就开始缓慢地重新计算，更不用说做公式、拉透视表、绘制分析图表了。

然后就是：在某个单元格输入一个数据，工作簿就立刻趴下“装死”，我们只好干瞪眼耐心地候着，就这样，半天的时间过去了，连一个最简单的同比分析报表都没做出来。

这种痛苦，这种逼疯人的节奏，只有使用者最有体会！

图1-3就是这样的一个示例，其数据有近80万行（见图1-4），保存为CSV格式的文本文件。