

LEARNING MADE EASY



5th Edition

# Microsoft® Excel® Data Analysis

for  
**dummies**®  
A Wiley Brand



Master Excel's powerful  
data-analysis tools

Generate data insights with  
PivotTables and trend forecasts

Visualize complex data with  
graphics and PivotCharts

**Paul McFedries**

Bestselling author of more than  
100 books



# Excel® Data Analysis

**for  
dummies®**  
A Wiley Brand





# Excel® Data Analysis

5th Edition

**by Paul McFedries**

**for  
dummies®**  
A Wiley Brand

## Excel® Data Analysis For Dummies®, 5th Edition

Published by: **John Wiley & Sons, Inc.**, 111 River Street, Hoboken, NJ 07030-5774, [www.wiley.com](http://www.wiley.com)

Copyright © 2022 by John Wiley & Sons, Inc., Hoboken, New Jersey

Media and software compilation copyright © 2022 by John Wiley & Sons, Inc. All rights reserved.

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

**Trademarks:** Wiley, For Dummies, the Dummies Man logo, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and may not be used without written permission. Microsoft and Excel are registered trademarks of Microsoft Corporation. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: WHILE THE PUBLISHER AND AUTHORS HAVE USED THEIR BEST EFFORTS IN PREPARING THIS WORK, THEY MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES REPRESENTATIVES, WRITTEN SALES MATERIALS OR PROMOTIONAL STATEMENTS FOR THIS WORK. THE FACT THAT AN ORGANIZATION, WEBSITE, OR PRODUCT IS REFERRED TO IN THIS WORK AS A CITATION AND/OR POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE PUBLISHER AND AUTHORS ENDORSE THE INFORMATION OR SERVICES THE ORGANIZATION, WEBSITE, OR PRODUCT MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING PROFESSIONAL SERVICES. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR YOUR SITUATION. YOU SHOULD CONSULT WITH A SPECIALIST WHERE APPROPRIATE. FURTHER, READERS SHOULD BE AWARE THAT WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ. NEITHER THE PUBLISHER NOR AUTHORS SHALL BE LIABLE FOR ANY LOSS OF PROFIT OR ANY OTHER COMMERCIAL DAMAGES, INCLUDING BUT NOT LIMITED TO SPECIAL, INCIDENTAL, CONSEQUENTIAL, OR OTHER DAMAGES.

For general information on our other products and services, please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993, or fax 317-572-4002. For technical support, please visit <https://hub.wiley.com/community/support/dummies>.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit [www.wiley.com](http://www.wiley.com).

Library of Congress Control Number is available from the publisher.

ISBN 978-1-119-84442-6 (pbk); ISBN 978-1-119-84446-4 (ebk); ISBN 978-1-119-84447-1 (ebk)

# Contents at a Glance

<b>Introduction</b> .....	1
<b>Part 1: Getting Started with Data Analysis</b> .....	5
CHAPTER 1: Learning Basic Data-Analysis Techniques .....	7
CHAPTER 2: Working with Data-Analysis Tools .....	31
CHAPTER 3: Introducing Excel Tables .....	55
CHAPTER 4: Grabbing Data from External Sources .....	77
CHAPTER 5: Analyzing Table Data with Functions .....	99
<b>Part 2: Analyzing Data Using PivotTables and PivotCharts</b> .....	113
CHAPTER 6: Creating and Using PivotTables .....	115
CHAPTER 7: Performing PivotTable Calculations .....	137
CHAPTER 8: Building PivotCharts .....	161
<b>Part 3: Discovering Advanced Data-Analysis Tools</b> .....	181
CHAPTER 9: Dealing with Data Models .....	183
CHAPTER 10: Tracking Trends and Making Forecasts .....	203
CHAPTER 11: Analyzing Data Using Statistics .....	227
CHAPTER 12: Analyzing Data Using Descriptive Statistics .....	245
CHAPTER 13: Analyzing Data Using Inferential Statistics .....	261
<b>Part 4: The Part of Tens</b> .....	279
CHAPTER 14: Ten Things You Ought to Know about Statistics .....	281
CHAPTER 15: Ten Ways to Analyze Financial Data .....	293
CHAPTER 16: Ten Ways to Raise Your PivotTable Game .....	303
<b>Appendix: Glossary of Data Analysis and Excel Terms</b> ....	317
<b>Index</b> .....	327





# Table of Contents

<b>INTRODUCTION</b>	<b>1</b>
About This Book	1
What You Can Safely Ignore	2
Foolish Assumptions	2
Icons Used in This Book	3
Beyond the Book	3
Where to Go from Here	4
<b>PART 1: GETTING STARTED WITH DATA ANALYSIS</b>	<b>5</b>
<b>CHAPTER 1: Learning Basic Data-Analysis Techniques</b>	<b>7</b>
What Is Data Analysis, Anyway?	8
Cooking raw data	8
Dealing with data	8
Building data models	9
Performing what-if analysis	9
Analyzing Data with Conditional Formatting	9
Highlighting cells that meet some criteria	10
Showing pesky duplicate values	11
Highlighting the top or bottom values in a range	13
Analyzing cell values with data bars	14
Analyzing cell values with color scales	15
Analyzing cell values with icon sets	16
Creating a custom conditional-formatting rule	17
Editing a conditional-formatting rule	20
Removing conditional-formatting rules	21
Summarizing Data with Subtotals	22
Grouping Related Data	24
Consolidating Data from Multiple Worksheets	26
Consolidating by position	26
Consolidating by category	28
<b>CHAPTER 2: Working with Data-Analysis Tools</b>	<b>31</b>
Working with Data Tables	32
Creating a basic data table	32
Creating a two-input data table	34
Skipping data tables when calculating workbooks	37
Analyzing Data with Goal Seek	37
Analyzing Data with Scenarios	39
Create a scenario	40
Apply a scenario	42

Edit a scenario .....	42
Delete a scenario .....	43
Optimizing Data with Solver .....	43
Understanding Solver .....	43
The advantages of Solver .....	44
When should you use Solver? .....	44
Loading the Solver add-in. ....	46
Optimizing a result with Solver .....	47
Adding constraints to Solver .....	50
Save a Solver solution as a scenario .....	52
<b>CHAPTER 3: Introducing Excel Tables .....</b>	<b>55</b>
What Is a Table and Why Should I Care? .....	55
Building a Table .....	57
Getting the data from an external source .....	58
Converting a range to a table .....	58
Basic table maintenance .....	60
Analyzing Table Information .....	61
Displaying simple statistics .....	61
Adding a column subtotal .....	63
Sorting table records. ....	64
Filtering table records. ....	67
Clearing a filter. ....	68
Turning off AutoFilter .....	68
Applying a predefined AutoFilter .....	69
Applying multiple filters .....	71
Applying advanced filters .....	72
<b>CHAPTER 4: Grabbing Data from External Sources .....</b>	<b>77</b>
What's All This about External Data? .....	77
Exporting Data from Other Programs .....	79
Importing External Data into Excel .....	80
Importing data from an Access table .....	80
Importing data from a Word table .....	81
Introducing text file importing. ....	82
Importing a delimited text file .....	83
Importing a fixed-width text file .....	84
Importing data from a web page .....	86
Importing an XML file .....	88
Querying External Databases .....	90
Defining a data source .....	91
Querying a data source .....	94

<b>CHAPTER 5: Analyzing Table Data with Functions</b>	99
The Database Functions: Some General Remarks	100
Retrieving a Value from a Table	101
Summing a Column's Values	102
Counting a Column's Values	104
Averaging a Column's Values	105
Determining a Column's Maximum and Minimum Values	106
Multiplying a Column's Values	108
Deriving a Column's Standard Deviation	109
Calculating a Column's Variance	110
 <b>PART 2: ANALYZING DATA USING PIVOTTABLES AND PIVOTCHARTS</b>	 113
<b>CHAPTER 6: Creating and Using PivotTables</b>	115
Understanding PivotTables	116
Exploring PivotTable Features	118
Building a PivotTable from an Excel Range or Table	119
Creating a PivotTable from External Data	122
Building a PivotTable from Microsoft Query	122
Building a PivotTable from a new data connection	123
Refreshing PivotTable Data	125
Refreshing PivotTable data manually	125
Refreshing PivotTable data automatically	125
Adding Multiple Fields to a PivotTable Area	126
Pivoting a Field to a Different Area	127
Grouping PivotTable Values	128
Grouping numeric values	128
Grouping date and time values	129
Grouping text values	130
Filtering PivotTable Values	131
Applying a report filter	131
Filtering row or column items	132
Filtering PivotTable values	133
Filtering a PivotTable with a slicer	134
 <b>CHAPTER 7: Performing PivotTable Calculations</b>	 137
Messing around with PivotTable Summary Calculations	138
Changing the PivotTable summary calculation	138
Trying out the difference summary calculation	140
Applying a percentage summary calculation	142
Adding a running total summary calculation	144
Creating an index summary calculation	147

Working with PivotTable Subtotals . . . . .	149
Turning off subtotals for a field . . . . .	149
Displaying multiple subtotals for a field . . . . .	149
Introducing Custom Calculations . . . . .	151
Formulas for custom calculations . . . . .	151
Checking out the custom calculation types . . . . .	152
Understanding custom calculation limitations . . . . .	152
Inserting a Custom Calculated Field . . . . .	154
Inserting a Custom Calculated Item . . . . .	155
Editing a Custom Calculation . . . . .	157
Deleting a Custom Calculation . . . . .	158
<b>CHAPTER 8: Building PivotCharts . . . . .</b>	<b>161</b>
Introducing the PivotChart . . . . .	161
Understanding PivotChart pros and cons . . . . .	162
Taking a PivotChart tour . . . . .	163
Understanding PivotChart limitations . . . . .	164
Creating a PivotChart . . . . .	165
Creating a PivotChart from a PivotTable . . . . .	165
Embedding a PivotChart on a PivotTable's worksheet . . . . .	165
Creating a PivotChart from an Excel range or table . . . . .	166
Working with PivotCharts . . . . .	169
Moving a PivotChart to another sheet . . . . .	169
Filtering a PivotChart . . . . .	170
Changing the PivotChart type . . . . .	172
Adding data labels to your PivotChart . . . . .	173
Sorting the PivotChart . . . . .	174
Adding PivotChart titles . . . . .	175
Moving the PivotChart legend . . . . .	177
Displaying a data table with the PivotChart . . . . .	178
<b>PART 3: DISCOVERING ADVANCED DATA-ANALYSIS TOOLS . . . . .</b>	<b>181</b>
<b>CHAPTER 9: Dealing with Data Models . . . . .</b>	<b>183</b>
Understanding Excel Data Models . . . . .	184
Creating a relationship between tables . . . . .	186
Importing related external data tables . . . . .	187
Basing a PivotTable on multiple, related tables . . . . .	189
Managing a Data Model with Power Pivot . . . . .	192
Enabling the Power Pivot add-in . . . . .	192
Adding a table to the Data Model . . . . .	193
Importing related tables from an external data source . . . . .	193
Viewing table relationships . . . . .	195

Viewing relationship details.....	196
Creating a relationship between tables with Power Pivot.....	197
Transforming Data .....	199
Creating a PivotTable or PivotChart from Your Data Model.....	200
<b>CHAPTER 10: Tracking Trends and Making Forecasts.....</b>	<b>203</b>
Plotting a Best-Fit Trend Line.....	204
Calculating Best-Fit Values.....	206
Plotting Forecasted Values.....	208
Extending a Linear Trend.....	210
Extending a linear trend using the fill handle.....	211
Extending a linear trend using the Series command.....	211
Calculating Forecasted Linear Values.....	212
Plotting an Exponential Trend Line.....	214
Calculating Exponential Trend Values .....	216
Plotting a Logarithmic Trend Line.....	218
Plotting a Power Trend Line.....	220
Plotting a Polynomial Trend Line .....	222
Creating a Forecast Sheet .....	223
<b>CHAPTER 11: Analyzing Data Using Statistics .....</b>	<b>227</b>
Counting Things.....	228
Counting numbers .....	228
Counting nonempty cells .....	229
Counting empty cells.....	229
Counting cells that match criteria.....	229
Counting cells that match multiple criteria .....	230
Counting permutations .....	231
Counting combinations .....	232
Averaging Things.....	233
Calculating an average .....	233
Calculating a conditional average.....	233
Calculating an average based on multiple conditions.....	234
Calculating the median.....	235
Calculating the mode .....	235
Finding the Rank .....	236
Determining the Nth Largest or Smallest Value .....	238
Calculating the nth highest value .....	238
Calculating the nth smallest value .....	239
Creating a Grouped Frequency Distribution.....	240
Calculating the Variance.....	241
Calculating the Standard Deviation .....	242
Finding the Correlation.....	243

<b>CHAPTER 12: Analyzing Data Using Descriptive Statistics</b>	245
Loading the Analysis ToolPak	246
Generating Descriptive Statistics	247
Calculating a Moving Average	250
Determining Rank and Percentile	253
Generating Random Numbers	255
Creating a Frequency Distribution	258
<b>CHAPTER 13: Analyzing Data Using Inferential Statistics</b>	261
Sampling Data	262
Using the t-Test Tools	264
Performing a z-Test	267
Determining the Regression	269
Calculating the Correlation	271
Calculating the Covariance	273
Using the Anova Tools	274
Performing an f-Test	276
<b>PART 4: THE PART OF TENS</b>	279
<b>CHAPTER 14: Ten Things You Ought to Know about Statistics</b>	281
Descriptive Statistics Are Straightforward	282
Averages Aren't So Simple Sometimes	282
Standard Deviations Describe Dispersion	283
An Observation Is an Observation	285
A Sample Is a Subset of Values	285
Inferential Statistics Are Cool But Complicated	285
Probability Distributions Aren't Always Confusing	287
Uniform distribution	287
Normal distribution	288
Parameters Aren't So Complicated	289
Skewness and Kurtosis Describe a Probability Distribution's Shape	289
Confidence Intervals Seem Complicated at First But Are Useful	291
<b>CHAPTER 15: Ten Ways to Analyze Financial Data</b>	293
Calculating Future Value	294
Calculating Present Value	295
Determining Loan Payments	296
Calculating a Loan Payment's Principal and Interest	297
Calculating Cumulative Loan Principal and Interest	297
Finding the Required Interest Rate	298
Determining the Internal Rate of Return	299

Calculating Straight-Line Depreciation.....	300
Returning the Fixed-Declining Balance Depreciation.....	300
Determining the Double-Declining Balance Depreciation.....	301
<b>CHAPTER 16: Ten Ways to Raise Your PivotTable Game.....</b>	<b>303</b>
Turn the PivotTable Fields Task Pane On and Off .....	304
Change the PivotTable Fields Task Pane Layout.....	304
Display the Details Behind PivotTable Data .....	306
Apply a PivotTable Style .....	308
Create a Custom PivotTable Style .....	309
Preserve PivotTable Formatting .....	311
Rename the PivotTable.....	312
Turn Off Grand Totals.....	312
Reduce the Size of PivotTable Workbooks.....	314
Use a PivotTable Value in a Formula .....	314
<b>APPENDIX: GLOSSARY OF DATA ANALYSIS</b>	
<b>AND EXCEL TERMS .....</b>	<b>317</b>
<b>INDEX.....</b>	<b>327</b>





# Introduction

---

**T**he world is bursting at the seams with data. It's on our computers, it's in our networks, it's on the web. Some days, it seems to be in the very air itself, borne on the wind. But here's the thing: No one actually cares about data. A collection of data — whether it resides on your PC or some giant server somewhere — is really just a bunch of numbers and text, dates and times. No one cares about data because data doesn't *mean* anything. Data isn't cool. You know what's cool? *Knowledge* is cool. *Insight* is cool.

So how do you turn data into knowledge? How do you tweak data to generate insight? You need to organize that data, and then you need to sort it, filter it, run calculations on it, and summarize it. In a word, you need to *analyze* the data.

Now for the good news: If you have (or can get) that data into Excel, you have a giant basket of data-analysis tools at your disposal. Excel really seems to have been made with data analysis in mind, because it offers such a wide variety of features and techniques for organizing, manipulating, and summarizing just about anything that resides in a worksheet. If you can get your data into Excel, it will help you turn that data into knowledge and insight.

This book takes you on a tour of Excel's data-analysis tools. You learn everything you need to know to make your data spill its secrets and to uncover your data's hidden-in-plain-sight wisdom. Best of all, if you already know how to perform the basic Excel chores, you don't need to learn any other fancy-schmancy Excel techniques to get started in data analysis. Sweet? You bet.

## About This Book

---

This book contains 16 chapters (and a bonus appendix), but that doesn't mean that you have to, as the King says gravely in *Alice's Adventures in Wonderland*, "Begin at the beginning and go on till you come to the end: Then stop." If you've done a bit of data-analysis work in the past, please feel free to dip into the book wherever it strikes your fancy. The chapters all present their data-analysis info and techniques in readily digestible, bite-sized chunks, so you can certainly graze your way through this book.

However, if you're brand spanking new to data analysis — particularly if you're not even sure what data analysis even is — no problem: I'm here to help. To get your data-analysis education off to a solid start, I highly recommend reading the book's first three chapters to get some of the basics down cold. From there, you can travel to more advanced territory, safe in the knowledge that you've got some survival skills to fall back on.

## What You Can Safely Ignore

This book consists of several hundred pages. Do I expect you to read every word on every page? Yes, I do. Just kidding! No, of course I don't. Entire sections — heck, maybe even entire *chapters* — might contain information that's not relevant to what you do. That's fine and my feelings won't be hurt if you skim through (or — who's kidding whom? — skip over) those parts of the book.

If time (or attention) is short, what else might you want to ignore? Okay, in many places throughout the book I provide step-by-step instructions to complete some task. Each of those steps includes some bold type that gives you the basic instruction. In many cases, however, below that bold text I offer supplementary information to flesh out or extend or explain the bold instruction. Am I just showing off how much I know about all this stuff? Yes, sometimes. Do you have to read these extended instructions? Nope. Read the bold stuff, for sure, but feel free to skip the details if they seem unnecessary or unimportant.

## Foolish Assumptions

This book is for people who are new (or relatively new) to Excel data analysis. That doesn't mean, however, that the book is suitable to people who have never used a PC, Microsoft Windows, or even Excel. So first I assume not only that you have a PC running Microsoft Windows but also that you've had some experience with both. (For the purposes of this book, that just means you know how to start and switch between programs.) I also assume that your PC has a recent version of Excel installed. What does "recent" mean? Well, this book is based on Excel 2021, but you should be fine if you're running Excel 365, Excel 2019, Excel 2016, or even Excel 2013.

As I said before, I do *not* assume that you're an Excel expert, but I do assume that you know at least the following Excel basics:

- » Creating, saving, opening, and switching between workbooks
- » Creating and switching between worksheets
- » Finding and running commands on the Ribbon
- » Entering numbers, text, dates, times, and formulas into worksheet cells
- » Working with Excel's basic worksheet functions

## Icons Used in This Book

Like other books in the *For Dummies* series, this book uses icons, or little margin pictures, to flag things that don't quite fit into the flow of the chapter discussion. Here are the icons that I use:



REMEMBER

This icon marks text that contains some things that are useful or important enough that you'd do well to store the text somewhere safe in your memory for later recall.



TECHNICAL  
STUFF

This icon marks text that contains some for-nerds-only technical details or explanations that you're free to skip.



TIP

This icon marks text that contains a shortcut or an easier way to do things, which I hope will make your life — or, at least, the data-analysis portion of your life — more efficient.



WARNING

This icon marks text that contains a friendly but unusually insistent reminder to avoid doing something. You have been warned.

## Beyond the Book

- » **Examples:** This book's sample Excel workbooks can be found by going to [www.dummies.com/go/exceldataanalysisfd5e](http://www.dummies.com/go/exceldataanalysisfd5e) or at my website: [www.paulmcfedries.com](http://www.paulmcfedries.com).
- » **Cheat Sheet:** To locate this book's cheat sheet, go to [www.dummies.com](http://www.dummies.com) and search for *Excel Data Analysis For Dummies*. See the cheat sheet for info on

Excel database functions, Boolean expressions, and important statistical terms.

» **Updates:** If this book has any updates after printing, they will be posted to this book's page at [www.dummies.com](http://www.dummies.com).

## Where to Go from Here

If you're just getting your feet wet with Excel data analysis, flip the page and start perusing the first chapter.

If you have some experience with Excel data analysis or you have a special problem or question, use the Table of Contents or the index to find out where I cover that topic and then turn to that page.

Either way, happy analyzing!

# 1

## **Getting Started with Data Analysis**

#### **IN THIS PART . . .**

Understand data analysis and get to know basic analysis features such as conditional formatting and subtotals.

Discover Excel's built-in tools for analyzing data.

Learn how to build Excel tables that hold and store the data you need to analyze.

Find quick and easy ways to begin your analysis using simple statistics, sorting, and filtering.

Get practical stratagems and common-sense tactics for grabbing data from extra sources.

- » Learning about data analysis
- » Analyzing data by applying conditional formatting
- » Adding subtotals to summarize data
- » Grouping related data
- » Combining data from multiple worksheets

## Chapter **1**

# Learning Basic Data-Analysis Techniques

**Y**ou are awash in data. Information multiplies around you so fast that you wonder how to make sense of it all. You think, “I know what to do. I’ll paste the data into Excel. That way, at least the data will be nicely arranged in the worksheet cells, and I can add a little formatting to make things somewhat palatable.” That’s a fine start, but you’re often called upon to do more with your data than make it merely presentable. Your boss, your customer, or perhaps just your curiosity requires you to divine some inner meaning from the jumble of numbers and text that litter your workbooks. In other words, you need to *analyze* your data to see what nuggets of understanding you can unearth.

This chapter gets you started down that data-analysis path by exploring a few straightforward but useful analytic techniques. After discovering what data analysis entails, you investigate a number of Excel data-analysis techniques, including conditional formatting, data bars, color scales, and icon sets. From there, you dive into some useful methods for summarizing your data, including subtotals, grouping, and consolidation. Before you know it, that untamed wilderness of a worksheet will be nicely groomed and landscaped.

# What Is Data Analysis, Anyway?

Are you wondering, “What is data analysis, anyway?” That’s an excellent question! Here’s an answer that I unpack for you as I go along: *Data analysis* is the application of tools and techniques to organize, study, reach conclusions, and sometimes make predictions about a specific collection of information.

For example, a sales manager might use data analysis to study the sales history of a product, determine the overall trend, and produce a forecast of future sales. A scientist might use data analysis to study experimental findings and determine the statistical significance of the results. A family might use data analysis to find the maximum mortgage it can afford or how much it must put aside each month to finance retirement or the kids’ education.

## Cooking raw data

The point of data analysis is to understand information on some deeper, more meaningful level. By definition, *raw data* is a mere collection of facts that by themselves tell you little or nothing of any importance. To gain some understanding of the data, you must manipulate the data in some meaningful way. The purpose of manipulating data can be something as simple as finding the sum or average of a column of numbers or as complex as employing a full-scale regression analysis to determine the underlying trend of a range of values. Both are examples of data analysis, and Excel offers a number of tools — from the straightforward to the sophisticated — to meet even the most demanding needs.

## Dealing with data

The *data* part of *data analysis* is a collection of numbers, dates, and text that represents the raw information you have to work with. In Excel, this data resides inside a worksheet, which makes the data available for you to apply Excel’s satisfyingly large array of data-analysis tools.

Most data-analysis projects involve large amounts of data, and the fastest and most accurate way to get that data onto a worksheet is to import it from a non-Excel data source. In the simplest scenario, you can copy the data from a text file, a Word table, or an Access datasheet and then paste it into a worksheet. However, most business and scientific data is stored in large databases, so Excel offers tools to import the data you need into your worksheet. I talk about all this in more detail later in the book.



After you have your data in the worksheet, you can use the data as is to apply many data-analysis techniques. However, if you convert the range into a *table*, Excel treats the data as a simple database and enables you to apply a number of database-specific analysis techniques to the table.

## Building data models

In many cases, you perform data analysis on worksheet values by organizing those values into a *data model*, a collection of cells designed as a worksheet version of some real-world concept or scenario. The model includes not only the raw data but also one or more cells that represent some analysis of the data. For example, a mortgage amortization model would have the mortgage data — interest rate, principal, and term — and cells that calculate the payment, principal, and interest over the term. For such calculations, you use formulas and Excel's built-in worksheet functions.

## Performing what-if analysis

One of the most common data-analysis techniques is *what-if analysis*, for which you set up worksheet models to analyze hypothetical situations. The “what-if” part means that these situations usually come in the form of a question: “What happens to the monthly payment if the interest rate goes up by 2 percent?” “What will the sales be if you increase the advertising budget by 10 percent?” Excel offers four what-if analysis tools: data tables, Goal Seek, Solver, and scenarios, all of which I cover in this book.

# Analyzing Data with Conditional Formatting

Many Excel worksheets contain hundreds of data values. You could try to make sense of such largish sets of data by creating complex formulas and wielding Excel's powerful data-analysis tools. However, just as you wouldn't use a steam-roller to crush a tin can, sometimes these sophisticated techniques are too much tool for the job. For example, what if all you want are answers to simple questions such as the following:

- » Which cell values are less than 0?
- » What are the top 10 values?
- » Which cell values are above average, and which are below average?

These simple questions aren't easy to answer just by glancing at the worksheet, and the more numbers you're dealing with, the harder it gets. To help you eyeball your worksheets and answer these and similar questions, Excel lets you apply conditional formatting to the cells. Excel applies this special format only to cells that satisfy some condition, which Excel calls a *rule*. For example, you could apply formatting to display all negative values in a red font, or you could apply a filter to show only the top 10 values.

## Highlighting cells that meet some criteria

A *conditional format* is formatting that Excel applies only to cells that meet the criteria you specify. For example, you can tell Excel to apply the formatting only if a cell's value is greater or less than some specified amount, between two specified values, or equal to some value. You can also look for cells that contain specified text, dates that occur during a specified time frame, and more.

When you set up your conditional format, you can specify the font, border, and background pattern. This formatting helps to ensure that the cells that meet your criteria stand out from the other cells in the range. Here are the steps to follow:

- 1. Select the range you want to work with.**

Select just the data values you want to format. Don't select any surrounding data.

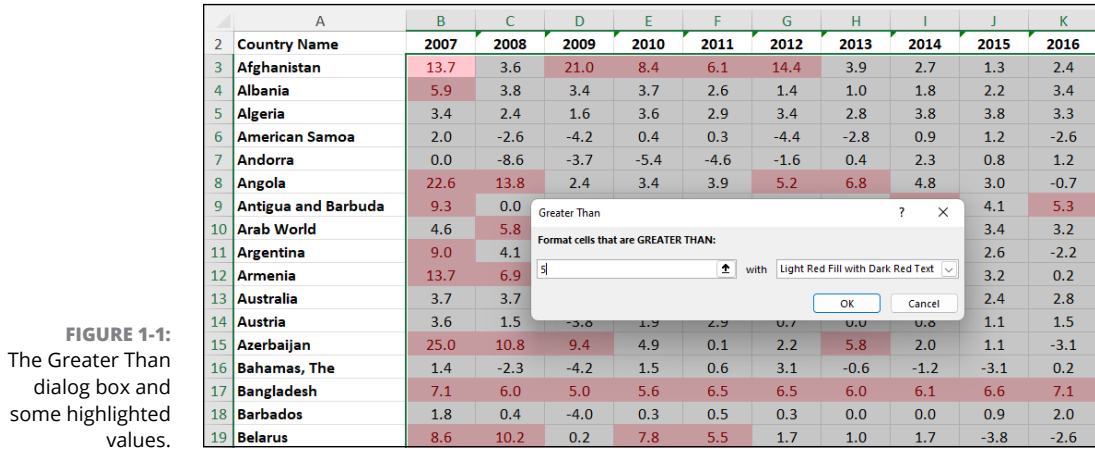
- 2. Choose Home ⇨ Conditional Formatting.**

- 3. Choose Highlight Cells Rules and then select the rule you want to use for the condition.**

You have six rules to play around with:

- **Greater Than:** Applies the conditional format to cells that have a value larger than a value that you specify.
- **Less Than:** Applies the conditional format to cells that have a value smaller than a value that you specify.
- **Between:** Applies the conditional format to cells that have a value that is greater than or equal to a minimum value that you specify and less than or equal to a maximum value that you specify.
- **Equal To:** Applies the conditional format to cells that have a value that is the same as a value that you specify.
- **Text that Contains:** Applies the conditional format to cells that include the text that you specify.
- **A Date Occurring:** Applies the conditional format to cells that have a date value that meets the condition that you specify (such as Yesterday, Last Week, or Next Month).

(I cover a seventh rule — Duplicate Values — later in this chapter.) A dialog box appears, the name of which depends on the rule you click in Step 3. For example, Figure 1-1 shows the dialog box for the Greater Than rule.



#### 4. Type the value to use for the condition.

You can also click the up arrow button that appears to the right of the text box and select a worksheet cell that contains the value. Also, depending on the operator, you might need to specify two values.

#### 5. Use the right drop-down list to select the formatting to apply to cells that match your condition.

If you're feeling creative, you can make up your own format by selecting the Custom Format command.

#### 6. Click OK.

Excel applies the formatting to cells that meet the condition you specified.



**TIP**

Excel enables you to specify multiple conditional formats for the same range. For example, you can set up one condition for cells that are greater than some value and a separate condition for cells that are less than some other value. You can apply unique formats to each condition. Keep the range selected and follow Steps 2 through 6 to configure the new condition.

## Showing pesky duplicate values

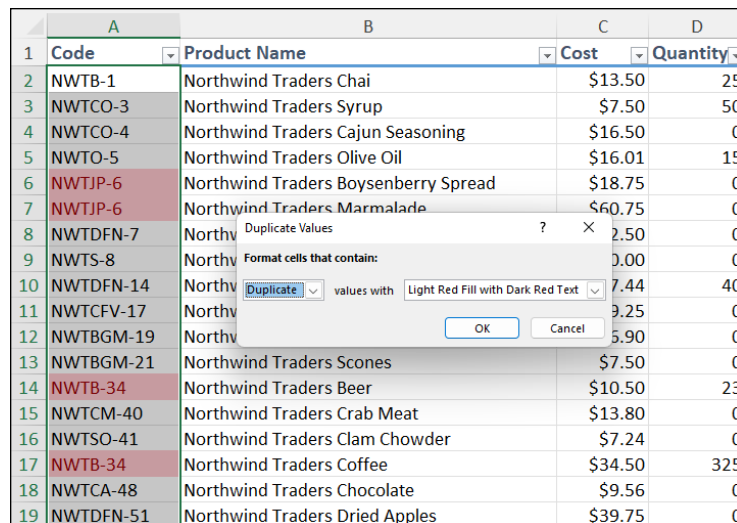
You use conditional formatting mostly to highlight numbers greater than or less than some value, or dates occurring within some range. However, you can use conditional formatting also to look for duplicate values in a range. Why would you

want to do that? The main reason is that many range or table columns require unique values. For example, a column of student IDs or part numbers shouldn't have duplicates.

Unfortunately, scanning such numbers and picking out the repeat values is hard. Not to worry! With conditional formatting, you can specify a font, border, and background pattern that ensures that any duplicate cells in a range or table stand out from the other cells. Here's what you do:

1. **Select the range that you want to check for duplicates.**
2. **Choose Home ⇨ Conditional Formatting.**
3. **Choose Highlight Cells Rules ⇨ Duplicate Values.**

The Duplicate Values dialog box appears. The left drop-down list has Duplicate selected by default, as shown in Figure 1-2. However, if you want to highlight all the unique values instead of the duplicates, select Unique from this list.



	A	B	C	D
1	Code	Product Name	Cost	Quantity
2	NWTB-1	Northwind Traders Chai	\$13.50	25
3	NWTCO-3	Northwind Traders Syrup	\$7.50	50
4	NWTCO-4	Northwind Traders Cajun Seasoning	\$16.50	0
5	NWTO-5	Northwind Traders Olive Oil	\$16.01	15
6	NWTJP-6	Northwind Traders Boysenberry Spread	\$18.75	0
7	NWTJP-6	Northwind Traders Marmalade	\$60.75	0
8	NWTDFN-7	Northwind Traders	2.50	0
9	NWTS-8	Northwind Traders	0.00	0
10	NWTDFN-14	Northwind Traders	7.44	40
11	NWTCFV-17	Northwind Traders	9.25	0
12	NWTBGM-19	Northwind Traders	5.90	0
13	NWTBGM-21	Northwind Traders Scones	\$7.50	0
14	NWTB-34	Northwind Traders Beer	\$10.50	23
15	NWTCM-40	Northwind Traders Crab Meat	\$13.80	0
16	NWTSO-41	Northwind Traders Clam Chowder	\$7.24	0
17	NWTB-34	Northwind Traders Coffee	\$34.50	325
18	NWTCA-48	Northwind Traders Chocolate	\$9.56	0
19	NWTDFN-51	Northwind Traders Dried Apples	\$39.75	0

**FIGURE 1-2:**  
Use the Duplicate Values rule to highlight worksheet duplicates.

4. **In the right drop-down list, select the formatting to apply to the cells with duplicate values.**

You can create your own format by choosing the Custom Format command. In the Format Cells dialog box, use the Font, Border, and Fill tabs to specify the formatting you want to apply, and then click OK.

5. **Click OK.**

Excel applies the formatting to any cells that have duplicate values in the range.

# Highlighting the top or bottom values in a range

When analyzing worksheet data, looking for items that stand out from the norm is often useful. For example, you might want to know which sales reps sold the most last year, or which departments had the lowest gross margins. To quickly and easily view the extreme values in a range, you can apply a conditional format to the top or bottom values of that range.

You can apply such a format by setting up a *top/bottom rule*, in which Excel applies a conditional format to those items that are at the top or bottom of a range of values. For the top or bottom values, you can specify a number, such as the top 5 or 10, or a percentage, such as the bottom 20 percent. Here's how it works:

1. **Select the range you want to work with.**
2. **Choose Home ⇨ Conditional Formatting.**
3. **Choose Top/Bottom Rules and then select the type of rule you want to create.**

You have six rules to mess with:

- **Top 10 Items:** Applies the conditional format to cells that rank in the top *X*, where *X* is a number that you specify (the default is 10).
- **Top 10 %:** Applies the conditional format to cells that rank in the top *X* %, where *X* is a number that you specify (the default is 10).
- **Bottom 10 Items:** Applies the conditional format to cells that rank in the bottom *X*, where *X* is a number that you specify (the default is 10).
- **Bottom 10 %:** Applies the conditional format to cells that rank in the bottom *X* %, where *X* is a number that you specify (the default is 10).
- **Above Average:** Applies the conditional format to cells that rank above the average value of the range.
- **Below Average:** Applies the conditional format to cells that rank below the average value of the range.

A dialog box appears, the name of which depends on the rule you selected in Step 3. For example, Figure 1-3 shows the dialog box for the Top Ten Items rule.

4. **Type the value to use for the condition.**

You can also click the spin buttons that appear to the right of the text box. Note that you don't need to enter a value for the Above Average and Below Average rules.

	A	B	C	D
1		Sales Rep	2017 Sales	2018 Sales
2		Nancy Freehafer	\$996,336	\$960,492
3		Andrew Cencini	\$606,731	\$577,983
4		Jan Kotas	\$622,781	\$967,580
5		Mariya Sergienko	\$765,327	\$771,399
6		Steven Thorpe	\$863,589	\$827,213
7		Michael Neinner	\$795,518	\$669,394
8			740	\$626,945
9			059	\$574,472
10			380	\$827,932
11			623	\$569,609
12			777	\$558,601
13		Paul Durbin	\$685,091	\$692,182
14		Andrea Granek	\$540,484	\$693,762
15		Charles Aster	\$650,733	\$823,034
16		Karen Reilly	\$509,863	\$511,569
17		Karen Munson	\$503,699	\$975,455
18		Vince Voyatzis	\$630,263	\$599,514
19		Paul Sellars	\$779,722	\$596,353

**FIGURE 1-3:**  
The Top 10 Items  
dialog box with  
the top 5 values  
highlighted.

- 5. In the right drop-down list, select the formatting to apply to cells that match your condition.**



**REMEMBER**

When you set up your top/bottom rule, select a format that ensures that the cells that meet your criteria will stand out from the other cells in the range. If none of the predefined formats suits your needs, you can always choose Custom Format and then use the Format Cells dialog box to create a suitable formatting combination. Use the Font, Border, and Fill tabs to specify the formatting you want to apply, and then click OK.

- 6. Click OK.**

Excel applies the formatting to cells that meet the condition you specified.

## Analyzing cell values with data bars

In some data-analysis scenarios, you might be interested more in the relative values within a range than the absolute values. For example, if you have a table of products that includes a column showing unit sales, you might want to compare the relative sales of all products.

Comparing relative values is often easiest if you visualize the values, and one of the easiest ways to visualize data in Excel is to use *data bars*, a data visualization feature that applies colored horizontal bars to each cell in a range of values; these bars appear “behind” (that is, in the background of) the values in the range. The length of the data bar in each cell depends on the value in that cell: the larger the value, the longer the data bar.