#### LEARNING MADE EASY



5th Edition

# Microsoft® Excel® Data Analysis



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

5th Edition

by Paul McFedries



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

Published by: John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, <u>www.wiley.com</u>

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 <u>https://hub.wiley.com/community/support/dummies</u>.

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 <u>http://booksupport.wiley.com</u>. For more information about Wiley products, visit <u>www.wiley.com</u>.

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)

### Excel® Data Analysis For Dummies®

To view this book's Cheat Sheet, simply go to <u>www.dummies.com</u> and search for "Excel Data Analysis For Dummies Cheat Sheet" in the Search box.

# **Table of Contents**

Cover **Title Page Copyright** Introduction About This Book What You Can Safely Ignore **Foolish Assumptions** Icons Used in This Book Beyond the Book Where to Go from Here Part 1: Getting Started with Data Analysis **Chapter 1: Learning Basic Data-Analysis Techniques** What Is Data Analysis, Anyway? Analyzing Data with Conditional Formatting Summarizing Data with Subtotals

#### Grouping Related Data

Consolidating Data from Multiple Worksheets

#### **Chapter 2: Working with Data-Analysis Tools**

Working with Data Tables

Analyzing Data with Goal Seek

Analyzing Data with Scenarios

**Optimizing Data with Solver** 

#### **Chapter 3: Introducing Excel Tables**

What Is a Table and Why Should I Care?

Building a Table

Analyzing Table Information

#### **Chapter 4: Grabbing Data from External Sources**

What's All This about External Data?

Exporting Data from Other Programs

Importing External Data into Excel

**Querying External Databases** 

#### **Chapter 5: Analyzing Table Data with Functions**

The Database Functions: Some General Remarks

Retrieving a Value from a Table

Summing a Column's Values

Counting a Column's Values

Averaging a Column's Values

Determining a Column's Maximum and Minimum Values

Multiplying a Column's Values

Deriving a Column's Standard Deviation

Calculating a Column's Variance

#### Part 2: Analyzing Data Using PivotTables and PivotCharts

**Chapter 6: Creating and Using PivotTables** 

Understanding PivotTables

Exploring PivotTable Features

Building a PivotTable from an Excel Range or Table

Creating a PivotTable from External Data

Refreshing PivotTable Data

Adding Multiple Fields to a PivotTable Area

Pivoting a Field to a Different Area

Grouping PivotTable Values

Filtering PivotTable Values

#### **Chapter 7: Performing PivotTable Calculations**

Messing around with PivotTable Summary Calculations

Working with PivotTable Subtotals

Introducing Custom Calculations

Inserting a Custom Calculated Field

Inserting a Custom Calculated Item

Editing a Custom Calculation

**Deleting a Custom Calculation** 

#### **Chapter 8: Building PivotCharts**

Introducing the PivotChart

Creating a PivotChart

Working with PivotCharts

#### Part 3: Discovering Advanced Data-Analysis Tools

#### **Chapter 9: Dealing with Data Models**

Understanding Excel Data Models
Managing a Data Model with Power Pivot

Transforming Data

Creating a PivotTable or PivotChart from Your Data Model

#### **Chapter 10: Tracking Trends and Making** Forecasts

Plotting a Best-Fit Trend Line

Calculating Best-Fit Values

**Plotting Forecasted Values** 

Extending a Linear Trend

Calculating Forecasted Linear Values

Plotting an Exponential Trend Line

Calculating Exponential Trend Values

Plotting a Logarithmic Trend Line

Plotting a Power Trend Line

Plotting a Polynomial Trend Line

Creating a Forecast Sheet

#### **Chapter 11: Analyzing Data Using Statistics**

**Counting Things** 

Averaging Things

Finding the Rank

Determining the Nth Largest or Smallest Value

Creating a Grouped Frequency Distribution

Calculating the Variance

Calculating the Standard Deviation

Finding the Correlation

#### Chapter 12: Analyzing Data Using Descriptive Statistics

- Loading the Analysis ToolPak
- **Generating Descriptive Statistics**

Calculating a Moving Average

**Determining Rank and Percentile** 

**Generating Random Numbers** 

Creating a Frequency Distribution

#### **Chapter 13: Analyzing Data Using Inferential Statistics**

<u>Sampling Data</u> <u>Using the t-Test Tools</u> <u>Performing a z-Test</u> <u>Determining the Regression</u> <u>Calculating the Correlation</u> <u>Calculating the Covariance</u> <u>Using the Anova Tools</u> Performing an f-Test

#### Part 4: The Part of Tens

#### Chapter 14: Ten Things You Ought to Know about Statistics

**Descriptive Statistics Are Straightforward** 

Averages Aren't So Simple Sometimes

Standard Deviations Describe Dispersion

An Observation Is an Observation

A Sample Is a Subset of Values

Inferential Statistics Are Cool But Complicated

Probability Distributions Aren't Always Confusing

Parameters Aren't So Complicated

<u>Skewness and Kurtosis Describe a Probability Distribution's</u> <u>Shape</u>

<u>Confidence Intervals Seem Complicated at First But Are</u> <u>Useful</u>

#### <u>Chapter 15: Ten Ways to Analyze Financial</u> <u>Data</u>

Calculating Future Value

Calculating Present Value

**Determining Loan Payments** 

Calculating a Loan Payment's Principal and Interest

Calculating Cumulative Loan Principal and Interest

Finding the Required Interest Rate

Determining the Internal Rate of Return

**Calculating Straight-Line Depreciation** 

Returning the Fixed-Declining Balance Depreciation

Determining the Double-Declining Balance Depreciation

<u>Chapter 16: Ten Ways to Raise Your</u> <u>PivotTable Game</u> Turn the PivotTable Fields Task Pane On and Off Change the PivotTable Fields Task Pane Layout Display the Details Behind PivotTable Data Apply a PivotTable Style Create a Custom PivotTable Style Preserve PivotTable Formatting Rename the PivotTable Turn Off Grand Totals Reduce the Size of PivotTable Workbooks Use a PivotTable Value in a Formula

#### <u>Appendix: Glossary of Data Analysis and Excel</u> <u>Terms</u>

<u>Index</u>

Author Bio

Advertisement Page

**Connect with Dummies** 

End User License Agreement

### **List of Tables**

#### Chapter 3

TABLE 3-1 Quick Statistical Measures Available on the Status Bar TABLE 3-2 Excel's Comparison Operators

#### **Chapter 12**

TABLE 12-1 Measures Generated by the Descriptive Statistics Tool TABLE 12-2 The Rank and Percentile Tool's Output Columns

### List of Illustrations

Chapter 1

FIGURE 1-1: The Greater Than dialog box and some highlighted values.

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

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

FIGURE 1-4: The higher the value, the longer the data bar.

FIGURE 1-5: Excel applies an icon based on each cell's value.

FIGURE 1-6: Use the New Formatting Rule dialog box to create a custom rule.

FIGURE 1-7: Use the Conditional Formatting Rules Manager to edit your rules.

FIGURE 1-8: Use the Subtotal dialog box to apply subtotals to a range.

FIGURE 1-9: Some subtotals applied to the Total column for each customer.

FIGURE 1-10: When you group a range, Excel displays its outlining tools.

FIGURE 1-11: Consolidate multiple worksheets by adding a range from each one.

FIGURE 1-12: When consolidating by category, tell Excel where your labels are I...

#### **Chapter 2**

FIGURE 2-1: This data table has the input values in a column.

FIGURE 2-2: Enter the address of the input cell.

FIGURE 2-3: The data table results.

FIGURE 2-4: For a two-input data table, enter one set of values in a column and...

FIGURE 2-5: Enter the addresses of the input cells.

FIGURE 2-6: The two-input data table results.

FIGURE 2-7: Using Goal Seek to calculate the annual deposit required to end up ...

FIGURE 2-8: Goal Seek took all of a second or two to find a solution.

FIGURE 2-9: Creating a scenario for a mortgage analysis.

FIGURE 2-10: Example values for a scenario's changing cells.

FIGURE 2-11: The goal for this data model is to find the breakeven point (wher...

FIGURE 2-12: The completed Solver Parameters dialog box.

FIGURE 2-13: The Solver Results dialog box and the solution to the break-even p...

FIGURE 2-14: The completed Add Constraint dialog box.

FIGURE 2-15: The Solver Results dialog box and the final solution to the break-...

#### **Chapter 3**

FIGURE 3-1: Some table terminology you should know.

FIGURE 3-2: Excel takes a stab at the full range you want to convert.

FIGURE 3-3: The range is now a full-fledged table, just like that.

FIGURE 3-4: Select a column's cells, and Excel displays a few stats in the stat...

FIGURE 3-5: Excel adds a Total row to the bottom of the table and displays the ...

FIGURE 3-6: Selecting the Quantity column's Sort & Filter button displays this ...

FIGURE 3-7: The Sort dialog box set up for a two-level sort.

FIGURE 3-8: Select the check box beside each column value that you want to incl...

FIGURE 3-9: The Parts table filtered to show only the records from Division 3.

FIGURE 3-10: The Custom AutoFilter dialog box.

FIGURE 3-11: The filter shown in Figure 3-10 applied to the Parts table.

FIGURE 3-12: A table set up for an advanced filter.

FIGURE 3-13: Put the comparison expressions on one row to match them all.

FIGURE 3-14: Put the comparison expressions on separate rows to match one or mo...

FIGURE 3-15: An advanced filter all set to go.

FIGURE 3-16: The results of the advanced filter in Figure 3-15.

#### **Chapter 4**

FIGURE 4-1: Select a table or query in the Navigator dialog box to see a previe...

FIGURE 4-2: Select the delimiter that gets the data into nice, neat columns.

FIGURE 4-3: If needed, edit the column starting points until your columns are c...

FIGURE 4-4: Choose the method you want to use to access the web data.

FIGURE 4-5: Select a table to preview its data.

FIGURE 4-6: Select an XML table to preview its data.

FIGURE 4-7: The Choose Data Source dialog box.

FIGURE 4-8: The completed Create New Data Source dialog box for the Northwind d...

FIGURE 4-9: Any data sources you've created appear in the Databases tab.

FIGURE 4-10: The completed Query Wizard — Choose Columns dialog box.

FIGURE 4-11: The Query Wizard – Filter Data dialog box with a filter added.

FIGURE 4-12: The Query Wizard — Sort Order dialog box.

#### **Chapter 5**

FIGURE 5-1: A table and its criteria range.

FIGURE 5-2: Use DGET to retrieve a value from a table based on your criteria.

FIGURE 5-3: Use DSUM to add a column's values based on your criteria.

FIGURE 5-4: Use DCOUNT (or DCOUNTA) to tally a column's values based on your cr...

FIGURE 5-5: Use DAVERAGE to average a column's values based on your criteria.

FIGURE 5-6: Use DMAX and DMIN to return a column's largest and smallest values ...

FIGURE 5-7: Use DPRODUCT to multiply a column's values based on your criteria.

FIGURE 5-8: Use DSTDEV (or DSTDEVP) to derive the standard deviation of a colum...

FIGURE 5-9: Use DVAR (or DVARP) to derive the variance of a column's values bas...

#### **Chapter 6**

FIGURE 6-1: Some great data, but how do you make sense of it?

FIGURE 6-2: The PivotTable creates order out of data chaos.

FIGURE 6-3: The features of a typical PivotTable.

FIGURE 6-4: You start with a blank PivotTable and the PivotTable Fields task pa...

FIGURE 6-5: The features of a typical PivotTable.

FIGURE 6-6: Import the external data to a PivotTable Report.

FIGURE 6-7: Select Refresh Data When Opening the File to tell Excel to refresh ...

FIGURE 6-8: The Grouping dialog box.

FIGURE 6-9: From the filter field's drop-down list, select a report filter.

FIGURE 6-10: Excel filters the PivotTable to show just the selected items in ea...

#### **Chapter 7**

FIGURE 7-1: Use the Value Field Settings dialog box to choose a summary calcula...

FIGURE 7-2: A PivotTable that shows sales in two years: 2021 and 2022.

FIGURE 7-3: Use the Value Field Settings dialog box to choose a summary calcula...

FIGURE 7-4: The PivotTable from Figure 7-2 is now using the Difference From cal...

FIGURE 7-5: A PivotTable that shows quarterly sales by region.

FIGURE 7-6: The PivotTable from Figure 7-5, now using the % Of calculation.

FIGURE 7-7: A PivotTable showing monthly order totals.

FIGURE 7-8: The PivotTable from Figure 7-7, with the Running Total In calculati...

FIGURE 7-9: A PivotTable showing units sold by category and region.

FIGURE 7-10: The PivotTable from Figure 7-9, with the Index calculation applied...

FIGURE 7-11: A PivotTable with multiple subtotals.

FIGURE 7-12: A custom calculated field, ready for insertion into the PivotTable...

FIGURE 7-13: The custom calculated field in action.

FIGURE 7-14: A custom calculated item, ready for action.

FIGURE 7-15: Two custom calculated items added to the row area.

#### **Chapter 8**

FIGURE 8-1: The major sights to see in the PivotChart landscape.

FIGURE 8-2: Excel kicks things off with a blank PivotTable and PivotChart, and ...

FIGURE 8-3: An embedded PivotChart and its PivotTable.

FIGURE 8-4: Use the Move Chart dialog box to move a PivotChart to another works...

FIGURE 8-5: Pull down the Report Filter field list and then select an item.

FIGURE 8-6: The PivotChart from Figure 8-5 now displayed as a pie chart.

FIGURE 8-7: A data table shown below a PivotChart.

#### **Chapter 9**

FIGURE 9-1: In the Orders table, the first record has an Order ID value of 1024...

FIGURE 9-2: In the Order Details table, the first three records have an Order I...

FIGURE 9-3: In the Create Relationship dialog box, relate your two tables by us...

FIGURE 9-4: You can click Select Related Tables to have Excel figure out which ...

FIGURE 9-5: When you import multiple, related tables, you can opt to create onl...

FIGURE 9-6: Be sure to select the Add This Data to the Data Model check box.

FIGURE 9-7: A PivotTable that summarizes data from two related tables: Orders a...

FIGURE 9-8: Excel displays this message if you try to add a field from an unrel...

FIGURE 9-9: Excel can automatically detect the required relationship.

FIGURE 9-10: Click Select Related Tables and Power Pivot will automatically sel...

FIGURE 9-11: Visualize your Data Model tables and relationships with diagram vi...

FIGURE 9-12: See the details of your Data Model relationships by using the Mana...

FIGURE 9-13: The Data Refresh dialog box shows you the progress of the refresh.

#### **Chapter 10**

FIGURE 10-1: A chart with a trend line and some geeky numbers.

FIGURE 10-2: An array of best-fit trend values (column F).

FIGURE 10-3: A best-fit trend line, extended four periods into the future.

FIGURE 10-4: The Series dialog box: your best-fit trend best friend.

FIGURE 10-5: An array of forecasted trend values (see the range D14:D17).

FIGURE 10-6: A chart with an exponential trend line.

FIGURE 10-7: An array of exponential trend values (column C) and forecast value...

FIGURE 10-8: A chart with a logarithmic trend line.

FIGURE 10-9: A chart with a power trend line.

FIGURE 10-10: A chart with a polynomial trend line.

FIGURE 10-11: You use the Create Forecast Worksheet dialog box to set up your f...

FIGURE 10-12: The new forecast worksheet showing a table and line chart for the...

#### Chapter 11

FIGURE 11-1: COUNT returns the tally of the numeric values in a range.

FIGURE 11-2: A parts database.

FIGURE 11-3: MEDIAN returns the median value of a set of numeric values.

FIGURE 11-4: RANK returns the ranking of a value in a set.

FIGURE 11-5: LARGE returns the *n*th largest value in a range or array.

FIGURE 11-6: FREQUENCY tells you how many items in a range appear in each bin.

FIGURE 11-7: VAR.P returns the variance of data that represents an entire popul...

FIGURE 11-8: CORREL calculates the correlation between two sets of values.

#### Chapter 12

FIGURE 12-1: Loading the Analysis ToolPak add-in.

FIGURE 12-2: The Analysis ToolPak's tools appear in the Data Analysis dialog bo...

FIGURE 12-3: The Descriptive Statistics dialog box, ready to run.

FIGURE 12-4: The statistics generated by the Descriptive Statistics tool.

FIGURE 12-5: The Moving Average dialog box, good to go.

FIGURE 12-6: Some moving averages generated by the Moving Average tool.

FIGURE 12-7: A completed version of the Rank and Percentile dialog box.

FIGURE 12-8: The product defects with their respective ranks and percentiles.

FIGURE 12-9: The parameters change depending on the Distribution value.

FIGURE 12-10: Random test scores generated by the Random Number Generation tool...

FIGURE 12-11: A worksheet with student grades and a bin range for the distribut...

FIGURE 12-12: The Histogram dialog box, all dressed up.

FIGURE 12-13: A frequency distribution and histogram for some student grades.

#### Chapter 13

FIGURE 13-1: A table of invoice data from which you might select a sample.

FIGURE 13-2: The Sampling dialog box, ready to sample.

FIGURE 13-3: The sample extracted by the Sampling tool.

FIGURE 13-4: Some sample data you can use to perform t-test calculations.

FIGURE 13-5: The t-Test: Two-Sample Assuming Equal Variances dialog box.

FIGURE 13-6: The results of a t-test.

FIGURE 13-7: The z-Test: Two Sample for Means dialog box.

FIGURE 13-8: The z-test calculation results.

FIGURE 13-9: The completed Regression dialog box.

FIGURE 13-10: The regression analysis results.

FIGURE 13-11: The completed Correlation dialog box.

FIGURE 13-12: The worksheet showing the correlation results for the list price ...

FIGURE 13-13: The completed Covariance dialog box.

FIGURE 13-14: The worksheet showing the covariance results for the list price a...

FIGURE 13-15: The completed Anova: Single Factor dialog box.

FIGURE 13-16: The worksheet showing the Anova results for the list price and un...

FIGURE 13-17: The completed F-Test Two-Sample for Variances dialog box.

FIGURE 13-18: The worksheet showing the F-Test Two-Sample for Variances results...

#### Chapter 14

FIGURE 14-1: A standard normal distribution.

FIGURE 14-2: A uniform distribution generated by 60,000 dice rolls.

FIGURE 14-3: Data sets that are negatively skewed (left) and positively skewed ...

FIGURE 14-4: Data sets that have negative kurtosis (left) and positive kurtosis...

#### **Chapter 16**

FIGURE 16-1: Click Tools to see the PivotTable Fields task pane layout options.

FIGURE 16-2: You've sold 792 earbuds, but what are the details behind this numb...

FIGURE 16-3: The details behind the earbud sales shown in Figure 16-2.

FIGURE 16-4: Use the New PivotTable Style dialog box to define your custom styl...

FIGURE 16-5: The GETPIVOTDATA function doing its thing.

# Introduction

The 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 dataanalysis info and techniques in readily digestible, bitesized 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 dataanalysis 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.



only technical details or explanations that you're free to skip.



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 <u>www.dummies.com/go/exceldataanalysisfd5e</u> or at my website: <u>www.paulmcfedries.com</u>.
- » Cheat Sheet: To locate this book's cheat sheet, go to 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 <u>www.dummies.com</u>.

## 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!

### Part 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.

### Chapter 1

# Learning Basic Data-Analysis Techniques

#### **IN THIS CHAPTER**

» Learning about data analysis

» Analyzing data by applying conditional formatting

» Adding subtotals to summarize data

- » Grouping related data
- » Combining data from multiple worksheets

You 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 steamroller 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?