LEARNING MADE EASY

Data Analytics & Visualization

ALL-IN-ONE

for dummies®

A Wiley Brand

6 Books in one!

Jack Hyman et al.

Data Analytics & Visualization

ALL-IN-ONE

by Jack Hyman; Luca Massaron; Paul McFedries; John Paul Mueller; Lillian Pierson; Jonathan Reichental, PhD; Joseph Schmuller; Alan Simon; and Allen G. Taylor

for dummies®
A Wiley Brand

Data Analytics & Visualization All-in-One For Dummies®

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

Published simultaneously in Canada

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.

Library of Congress Control Number: 2024932207

ISBN 978-1-394-24409-6 (pbk); ISBN 978-1-394-24411-9 (ePDF); ISBN 978-1-394-24410-2 (epub)

Contents at a Glance

Table of Contents

Introduction

Everywhere you go in the business world, you are likely to encounter executives who make decisions driven by tidbits of raw data that together tell a meaningful story. In fact, in our everyday worlds, websites and mobile apps express data using powerful visualizations to explain complex numbers and concepts, not extensive written passages anymore. The phrase "a picture speaks a thousand words" rings true in the world of data analytics and visualization, and for good reason.

Data analytics and visualization allow anyone to turn raw data into meaningful stories and insights. You, as the analyst, act as the detective. Instead of having to solve a mystery with clues, you are provided datasets that, if provided with enough clarity, can answer complex questions using trend and pattern analysis. If you review a dataset enough, you'll inevitably have an ah-ha moment in your interpretation quest, but if the dataset can be presented visually, you can accelerate your understanding like a racecar going from 0 to 100 miles per hour in seconds.

Data analytics and visualization help you uncover creative ways to showcase data in a manner that is both informative and engaging. Data often starts out as nothing more than a bunch of jumbled numbers; turning those numbers into a story that can influence decisions and drive change is incredibly powerful. Global enterprises rely on folks who have the skills you are about to embark on in this book as a way to determine business strategies, make corporate decisions, and influence change. If you are ready to learn these skills, you are in for a treat with this book.

About This Book

If you've picked up this book, you might be on a quest to piece together a whole lot of terms being thrown around in the information economy regarding data, the most precious tool in the information economy. Data is a business asset that sits at the intersection of many disciplines; the resultant product from data can be methodologies, processes, algorithms, and system outputs. To the end user though, the end game is extracting knowledge and insights from the byproducts of data, and taking action upon review.

Book 1 covers the foundational aspects of the data analytics and visualization lifecycle that every user must understand to be proficient as an analytics and visualization savvy. Books 2 and 3 focus on the two leading tools in the enterprise business intelligence market used to perform complex data analytics and visualization tasks; Microsoft Power BI and Tableau. Books 4 through 6 cover the key programming languages used by both proprietary and open-source data analytics and visualization platforms to extract, assess, and visualize data at scale when commercial off-the-shelf enterprise business platforms are unavailable.

This book uses the following technical conventions:

- Bold text means that you're meant to type the text just as it appears in the book. The exception is when you're working through a steps list: Because each step is bold, the text to type is not bold.
- Web addresses and programming code appear in monofont. If you're reading a digital version of this book on a device connected to the Internet, note that you can click the web address to visit that website, like this: `www.dummies.com`.
- For command sequences in software, this book uses the command arrow. Here's an example that uses Microsoft Word: Click the Office button and then choose Page Layout➪Margins➪Narrow to decrease the default margin setting.

If you don't think the book contains any conventions that need to be spelled out in this section, discuss omitting conventions information with your editor.

To make the content more accessible, we divided it into 6 books:

- **Book 1, "Learning Data Analytics & Visualization Foundations."**

  Book 1 introduces terms and fundamental concepts. You learn about big data, data lakes, and data science, and you see how you can apply visualization tools to create meaningful stories based on data you collect.

- **Book 2, "Using Power BI for Data Analysis & Visualization."**

  Book 2 covers Microsoft Power BI, a data analysis and visualization tool used by many large organizations. This book illustrates how you can use Power BI to make sense of structured, unstructured, and semi-structured data, and develop robust business analytics outputs for your organization.

- **Book 3, "Using Tableau for Data Analysis & Visualization."**

  Book 3 covers Tableau, a data analysis and visualization tool favored by researchers and educational institutions. In this book, you discover how to prepare data and present your findings using Tableau's storytelling and visualization features. You also see how to collaborate and publish your work with Tableau Cloud.

- **Book 4, "Extracting Information with SQL."**

  Book 4 describes SQL and the relational database model. You discover how SQL is a powerful tool that nonprogrammers can use to write complex queries to get the most out of their data, and more.

- **Book 5, "Performing Statistical Data Analysis & Visualization with R Programming."**

  Book 5 introduces the open-source R programming language. You see how you can use R to perform statistical data analysis, data visualization, and other data science tasks.

- **Book 6, "Applying Python Programming to Data Science."**

  Book 6 describes how Python is used as a data science and visualization tool. The book includes a "crash course" on MatPlotLib.

Foolish Assumptions

To get the most out of this book, you need the following:

- **Access to the Internet:** This may sound a bit obvious. Even with the Desktop client, an Internet connection is required in order to access datasets from the Internet.
- **A meaningful dataset:** A meaningful dataset includes at least 300 to 400 records containing a minimum of five or six columns' worth of data.

Icons Used in This Book

Throughout this book, icons in the margins highlight certain types of valuable information that call out for your attention. Here are the icons you'll encounter and a brief description of each.

BEST PRACTICE

Best Practice icons highlight points of common knowledge among seasoned professionals in the data industry. If you don't want to look like a complete newbie, follow the well-worn advice described in these paragraphs.

TIP

Tips point out shortcuts or essential suggestions that you can use to do things quicker, faster, and more efficiently.

REMEMBER

Consider these small suggestions that are quite helpful. Remember icons are like signs on the road to suggest a potential better route.

TECHNICAL STUFF

The Technical Stuff icon marks information of a highly technical nature that you can normally skip over. When appropriate, these paragraphs also suggest specialized resources you may find helpful down the road.

WARNING

The Warning icon makes you aware of a common issue or product challenge many users face. Don't fret, but *do* take note when you see this icon.

Beyond the Book

In addition to the abundance of information and guidance related to data analysis and visualization provided in this book, you get access to even more help and information online at ***Dummies.com***. Check out this book's online Cheat Sheet. Just go to ***www.dummies.com*** and search for "Data Analysis & Visualization All-in-One For Dummies Cheat Sheet."

Where to Go from Here

The book has three core themes: foundational concepts, tools, and programming languages.

If you want to learn the essential data analytics and visualization concepts, including learning the lingo of the land, head to Book 1.

If you're looking to get up to speed on Microsoft's Enterprise BI tools, head to Book 2. Tableau, a tool used for Enterprise BI but heavily leveraged in communities where data is regulated such as banking, healthcare, insurance, and government, head to Book 3.

The underpinning for data analytics and visualization is SQL, a querying language. To get a crash course on SQL, which is necessary for any proprietary or open-source data analytics and visualization platform, head to Book 4.

Finally, Books 5 and 6 are an introduction to two popular open-source programming languages, R and Python. Both languages can be configured for use with Power BI and Tableau, but are more commonly used with open-source (free) platforms like Jupyter Notebook and Anaconda to conceive data analytics outputs and visualizations. Unlike Power BI and Tableau, open-source tools leveraging programming languages are used in academic settings or by analysts requiring technologies that are data intensive.

1 Learning Data Analytics & Visualizations Foundations

Contents at a Glance