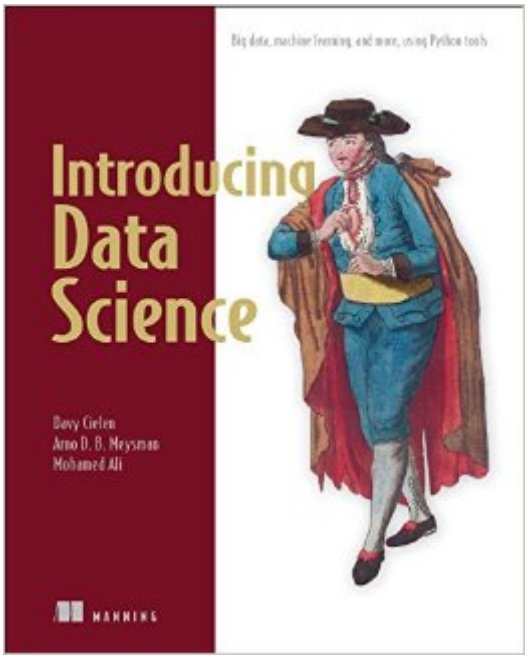
# Introducing Data Science: Big Data, Machine Learning, And More, Using Python Tools



## Synopsis

Summary Introducing Data Science teaches you how to accomplish the fundamental tasks that occupy data scientists. Using the Python language and common Python libraries, you'll experience firsthand the challenges of dealing with data at scale and gain a solid foundation in data science. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications.About the TechnologyMany companies need developers with data science skills to work on projects ranging from social media marketing to machine learning. Discovering what you need to learn to begin a career as a data scientist can seem bewildering. This book is designed to help you get started.About the BookIntroducing Data ScienceIntroducing Data Science explains vital data science concepts and teaches you how to accomplish the fundamental tasks that occupy data scientists. Youâ ™ll explore data visualization, graph databases, the use of NoSQL, and the data science process. Youâ ™ll use the Python language and common Python libraries as you experience firsthand the challenges of dealing with data at scale. Discover how Python allows you to gain insights from data sets so big that they need to be stored on multiple machines, or from data moving so quickly that no single machine can handle it. This book gives you hands-on experience with the most popular Python data science libraries, Scikit-learn and StatsModels. After reading this book, youâ ™ll have the solid foundation you need to start a career in data science. Whatâ ™s InsideHandling large dataIntroduction to machine learningUsing Python to work with dataWriting data science algorithmsAbout the ReaderThis book assumes you're comfortable reading code in Python or a similar language, such as C, Ruby, or JavaScript. No prior experience with data science is required.About the AuthorsDavy Cielen, Arno D. B. Meysman, and Mohamed Ali are the founders and managing partners of Optimately and Maiton, where they focus on developing data science projects and solutions in various sectors. Table of ContentsData science in a big data worldThe data science processMachine learningHandling large data on a single computerFirst steps in big dataJoin the NoSQL movementThe rise of graph databasesText mining and text analyticsData visualization to the end user

## Book Information

Paperback: 320 pages
Publisher: Manning Publications; 1 edition (May 23, 2016)
Language: English
ISBN-10: 1633430030
ISBN-13: 978-1633430037

Product Dimensions:  7.3 x 0.4 x 9.2 inches

Shipping Weight: 1.3 pounds (View shipping rates and policies)

Average Customer Review:  4.0 out of 5 starsÂ Â See all reviewsÂ (4 customer reviews)

Best Sellers Rank: #422,330 in Books (See Top 100 in Books)  #28 inÂ Books > Computers & Technology > Networking & Cloud Computing > Data in the Enterprise > Electronic Data Interchange (EDI)  #72 inÂ Books > Computers & Technology > Computer Science > AI & Machine Learning > Machine Theory  #74 inÂ Books > Computers & Technology > Programming > Languages & Tools > Ruby

## Customer Reviews

Loved this book! If I could have given 6 stars, I would have.This book would provide you with a very well rounded approach to Data Science and by that I mean truly would give you a ride though all the aspects of this field versus showing you some regression algorithm using python and call it Data Science.Book has it all - not only it leverages probably the most favorite language (python) for its examples, it also goes in details in supporting tools and eco systems. For examples, Spark - Why create something when Spark is already here and we can just use it in our work.It covered NoSQL technologies to give readers enough information to get started and weighted pros and cons of each. I especially enjoyed reading ACID, BASE and CAP theorem sections. I am familiar with them and gave presentation on exact same topic few years ago and I enjoyed the read since it covered the important key points leaving me with nice warm feeling in my stomach that unaware readers will be in a good hands!During discussion of NoSQL, ElasticSearch was introduced and entire chapter was devoted on how to leverage search capabilities to provide us with valuable results... Search is something that ElasticSearch does best! Section about Damerau-Levenshtein was great. It made you think of dirty data that is present in the real world and how you deal with it (vs giving you example with perfectly clean and ready to use data)Speaking of real world experience - this book took a step back and instead of trying to be data science book and throwing cool python libraries at you, it talked about general approach in the real word when you deal with data science projects by trying to make you think of project’s research goals - Why are we doing this?

â œData science,â • the three authors of this book point out, â œis a very wide field, so wide indeed that a book ten times the size of this one wouldnâ ™t be able to cover it all. For each chapter, we picked a different aspect we find interesting. Some hard decisions had to be made to keep this book from collapsing your bookshelf!â •In my view, they have made very good choices. This

â œIntroducingâ • book is written well and logically organized. And it generally is aimed at individual computer users and persons contemplating possible careers in data science. The book also could be good for managers and others trying to get a handle on how some data science techniques could be brought to bear on their growing mounds of business data.If you are impatient to dive straight into dicing, slicing and graphing big data, you should know that books from Manning generally donâ ™t follow that kind of quick approach. You get some overviews, explanations and theory first, and then you ease into the heart of the matter. In this book, you get to â œFirst steps in big dataâ • in chapter five, after first delving into the data science process: 1. Setting the research goal; 2. Retrieving data; 3. Data preparation, 4. Data exploration; 5. Data modeling; and 6. Presentation and automation.Chapter five also is preceded by chapters on machine learning and how to handle large data files on a single computer.The â œFirst stepsâ • chapter, meanwhile, shows how to work (at the sandbox level) with two big data applications, Hadoop and Spark, and demonstrates how Python can be used to write big data jobs.

code, Database) Python Data Analytics: Data Analysis and Science using pandas, matplotlib and the Python Programming Language Learning: 25 Learning Techniques for Accelerated Learning - Learn Faster by 300%! (Learning, Memory Techniques, Accelerated Learning, Memory, E Learning, ... Learning Techniques, Exam Preparation) Big Data For Beginners: Understanding SMART Big Data, Data Mining & Data Analytics For improved Business Performance, Life Decisions & More! Bread Machine Cookbook: 101 Delicious, Nutritious, Low Budget, Mouthwatering Bread Machine Cookbook: Best Bread Machine Bread Recipe Recipes for Perfect-Every-Time Bread-From Every Kind of Machine Python: Learn Python in One Day and Learn It Well. Python for Beginners with Hands-on Project. (Learn Coding Fast with Hands-On Project Book 1) Programming Raspberry Pi 3: Getting Started With Python (Programming Raspberry Pi 3, Raspberry Pi 3 User Guide, Python Programming, Raspberry Pi 3 with Python Programming) Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault Learn: Cognitive Psychology - How to Learn, Any Skill or Subject in 21 Days! (Learn, Learning Disability, Learning Games, Learning Techniques, Learning ... Learning, Cognitive Science, Study) Python: A Beginner to Expert Guide to Learning the basics of Python Programming (Computer Science Series) Data Analytics: Practical Data Analysis and Statistical Guide to Transform and Evolve Any Business Leveraging the Power of Data Analytics, Data Science, ... (Hacking Freedom and Data Driven Book 2)